

**EVALUAR LA PERTINENCIA DE INTRODUCIR
ÁREAS TERMINALES DE ANÁLISIS DE DATOS EN
LA CARRERA EN FÍSICA**

TESIS

PRESENTADA AL

COLEGIO DE FÍSICA

COMO REQUISITO PARA OBTENER EL GRADO DE

LICENCIADO EN FÍSICA

POR

YAHIEL REYES MUNGUÍA

ASESORADO POR

DR. JORGE VELÁZQUEZ CASTRO



BUAP

PUEBLA, PUE. 2025

Resumen

Existe una creciente necesidad de modernizar los programas educativos para responder a las demandas del mercado laboral, el cual ha sido transformado por el auge de tecnologías digitales y el uso intensivo de datos. Estas transformaciones han incrementado significativamente la demanda de habilidades en análisis de datos en múltiples sectores, particularmente en aquellos vinculados con la ciencia, la tecnología, la ingeniería y las matemáticas, donde los egresados con estas competencias son cada vez más requeridos por su capacidad para interpretar información compleja y apoyar en la toma de decisiones informadas. Esta investigación se ha centrado en evaluar la pertinencia de integrar estas habilidades en el currículo de la carrera de física, considerando tanto las tendencias laborales actuales como las percepciones de estudiantes y empleadores utilizando metodologías que combinan análisis de datos secundarios provenientes de fuentes confiables, como la Encuesta Nacional de Ocupación y Empleo (ENOE) y el Instituto Mexicano para la Competitividad (IMCO), con datos primarios obtenidos a través de encuestas dirigidas a estudiantes de la Facultad de Ciencias Físico-Matemáticas de la Benemérita Universidad Autónoma de Puebla. A partir de un análisis detallado de datos ocupacionales, encuestas a estudiantes y una revisión de tendencias en el mercado laboral se evalúa la necesidad de fortalecer las competencias en análisis de datos entre los futuros egresados de la Licenciatura en Física.

“El éxito en la ciencia no es tanto una cuestión de logros individuales, sino del proceso de aprendizaje y descubrimiento mismo”.

— Carl Sagan.

Reconocimientos

Es difícil conocer a la ciencia si nunca entras realmente en contacto con ella.

A mis padres: ustedes procuraron siempre que desde pequeño fuera un niño que tuviera mucha curiosidad y me alentaban a buscar las respuestas para las mil preguntas que yo tenía.

Gracias mamá y papá por animarme y darme las herramientas, recursos y conocimientos necesarios para siempre poder seguir adelante. Estoy eternamente agradecido por todo su apoyo e influencia hacia el camino profesional que decidí escoger, no habría sido posible tomarlo sin ustedes.

A mi hermano Gengis, quien también contribuyó decisivamente a marcar el rumbo de mi trayectoria académica al haber recorrido el camino de la ingeniería y explicarme las diferencias que había, de forma que descubrí a la Física como licenciatura la cual encajaba perfectamente con mis intereses y aspiraciones. Con sus consejos pude embarcarme con plena confianza y un claro sentido del propósito en esa carrera.

Quiero agradecerle también a la Dra. Isabel Pedraza Morales por todo lo que hacía para motivarnos en la universidad mientras era nuestra tutora, por todos esos chocolates suizos, souvenirs del CERN e incluso por pagar de su propio bolsillo cursos de programación para nosotros, así como incluirnos en proyectos de alto renombre con sus propios

colegas del trabajo. Fue por ella que pude conocer cómo era la Física de frontera en el MIT y en muchas ocasiones eso fue un motivo muy presente para recordarme por qué seguir avanzando.

Por último, agradezco mucho a mi asesor de tesis, el Dr. Jorge Velázquez Castro por recibirme aquel día, escuchar mi historia y aceptar ser mi tutor para el tema de tesis que me apasionaba, su calidez, empatía y vasta experiencia fueron también un gran soporte para no perder el camino.

Gracias por toda su guía.

A mi familia, quienes siempre me animaron a cuestionar.
Y a todos los que me hicieron ser una mejor persona.

Contenido

Resumen	ii
Reconocimientos	iv
Capítulo 1	9
Introducción	9
1.1 Justificaciones.....	9
1.2 Objetivos	10
1.3 Motivación.....	11
1.4 Alcances y limitaciones	13
Capítulo 2	14
Marco Referencial.....	14
2.1 Análisis de datos en México.....	14
2.2 Personas e instituciones.....	14
2.3 Contexto y utilidad	17
2.4 Últimos trabajos y publicaciones	19
2.4.1 Salud y COVID-19	19
2.4.2 Economía y Finanzas.....	21
2.4.3 Ciencias de la Tierra y Geología.....	22
2.4.4 Educación y Tecnología	22
Capítulo 3	23
Marco Contextual	23
3.1 Ubicación espacial y geográfica	23
3.2 La academia en la industria.....	25
3.3 La explosión de datos en la industria.....	26
Capítulo 4	29
Marco Teórico.....	29
4.1 Análisis de datos	29

4.2	Negocios	31
4.3	Patrones.....	32
4.4	Inteligencia de negocios (BI)	32
4.5	Dato	33
4.5.1	Base de datos.....	35
4.5.2	Almacén de datos.....	36
4.6	Minería de datos.....	37
4.6.1	Técnicas de minería de datos	38
Capítulo 5		41
<hr/>		
	Metodología.....	41
5.1	Diseño de la investigación.....	42
5.2	Sujetos de estudio.....	43
5.3	Técnicas de investigación	44
5.4	Instrumentos de recolección de datos.....	44
5.5	Procedimientos	45
5.6	Variables e indicadores	46
5.7	Encuesta.....	48
5.8	Análisis de datos	50
Capítulo 6		52
<hr/>		
	Resultados.....	52
6.1	Aplicación de la encuesta.....	52
6.2	Limpieza de datos del repositorio para análisis de datos	54
6.3	Distribución por área de estudio	57
6.4	Prueba χ^2	61
6.5	Ingresos por percentiles y promedio.....	63
6.6	Retorno Sobre la Inversión y Acceso a Empleo de Calidad.....	68
	Conclusiones.....	72
	Tabla de contenidos	74
	Referencias	75

Capítulo 1

Introducción

1.1 Justificaciones

El análisis de datos es una herramienta indispensable para las ciencias exactas, ya que permite a los científicos y estudiantes comprender mejor los patrones y tendencias en los datos recopilados, lo que puede llevar a descubrimientos importantes y a la toma de decisiones informadas. Las áreas terminales en análisis de datos contribuyen a una comprensión profunda de los métodos y técnicas utilizados en la manipulación y visualización de datos, que son esenciales para la investigación y el desarrollo en muchas ciencias y disciplinas.

Así mismo, dichas áreas pueden ayudar a ampliar el campo laboral de los estudiantes de ciencias exactas de varias maneras:

Habilidades en demanda: como habilidad altamente valorada en muchos campos, incluyendo la tecnología, las finanzas, la salud y la investigación científica. Al poseer habilidades en análisis de datos, los estudiantes se destacan en el mercado laboral.

Mejora la investigación: como parte crítica de muchas investigaciones científicas, y al tener una comprensión profunda de los métodos y técnicas de análisis de datos, los estudiantes pueden realizar investigaciones más efectivas y rigurosas.

Nuevas oportunidades de trabajo: brindando a los estudiantes la oportunidad de explorar nuevos campos y puestos altamente demandados, como el de data scientist, data analyst o data engineer.

Mejora la toma de decisiones: Al tener una comprensión bien cimentada de cómo analizar y visualizar datos, los estudiantes estarán mejor preparados para tomar decisiones informadas a lo largo de sus carreras y en la sociedad en general.

1.2 Objetivos

El objetivo general de este trabajo es evaluar la pertinencia de incorporar áreas terminales de análisis de datos en el currículo de la carrera de Física. Se investigará de qué manera esta integración podría ofrecer beneficios académicos y profesionales a los estudiantes, con el fin de determinar si la inclusión de estas competencias en su plan de estudios tiene un impacto significativo en su formación académica y preparación profesional.

En concreto, se buscará responder la siguiente pregunta de investigación general:

- ¿Qué beneficios académicos y profesionales podrían obtenerse de la inclusión de áreas terminales de análisis de datos en el programa de estudios de Física?

Además, se abordarán las siguientes preguntas particulares:

- ¿Cómo puede la integración de áreas de análisis de datos en el currículo de física mejorar la empleabilidad y las oportunidades de investigación para los graduados?
- ¿Cuál es la percepción de los estudiantes y egresados de la carrera de física sobre la inclusión de cursos de análisis de datos en su formación académica?
- ¿Qué competencias específicas en análisis de datos consideran más relevantes los empleadores en sectores relacionados con la física?

Para evaluar el impacto de esta integración en la formación de los estudiantes, la investigación se enfocará en:

- Analizar cómo la inclusión de áreas de análisis de datos en el currículo de la carrera de física puede mejorar la empleabilidad y las oportunidades de investigación para los graduados.
- Describir la percepción de los estudiantes y egresados de física sobre la inclusión de cursos de análisis de datos en su formación académica.
- Indagar cuáles competencias involucradas en el análisis de datos son valoradas por sectores laborales interesados en los egresados de física.

1.3 Motivación

Los campos académicos de la ciencia, tecnología, ingeniería y matemáticas (conocidos en inglés como STEM) están creciendo y los trabajos relacionados en el área tienen una demanda cada vez mayor principalmente en países desarrollados como los Estados Unidos. Los empleadores buscan profesionales con habilidades comprobadas en ciencia, tecnología, ingeniería y matemáticas para impulsar la innovación y aprovechar al máximo sus inversiones [1].

Desafortunadamente para ellos, y afortunadamente para los estudiantes de estas áreas, la mayoría de los profesionales STEM están fuera del mercado laboral y ya se encuentran trabajando, en cambio, la oferta de talento STEM disponible se está agotando mientras que el mercado laboral se inunda cada vez más de trabajos STEM disponibles, lo cual resulta en una verdadera oportunidad para aquellos graduados en las áreas pertinentes [2].

De acuerdo con el Censo de Población y Vivienda 2020 elaborado por el Instituto Nacional de Estadística y Geografía (INEGI), la distribución de la población de 18 años y más con educación superior por campo de formación académica, de 19.5 millones de la población mexicana con educación superior sólo el 2.6% y el 16.9% estudió en el campo de formación de ciencias naturales, matemáticas y estadística,

e ingeniería, manufactura y construcción respectivamente [3]. Asimismo, el Departamento de Comercio de los Estados Unidos reconoce que las profesiones en las áreas STEM son algunas de las más remuneradas, las cuales poseen un mayor potencial de desarrollo profesional desde inicios del siglo XXI. Con datos del Centro de Investigaciones Pew, un trabajador promedio STEM gana dos tercios más que los empleados en otros campos [4].

Una investigación acerca de la situación laboral más reciente de los egresados de la Licenciatura en Física e Ingeniería Física en México que se llevó a cabo por parte de la Revista Mexicana de Física muestra que el 31.4% y el 22.5% de los egresados de las Licenciaturas en Física, Ingeniería Física y afines escogen la docencia y la investigación respectivamente como primer empleo, mientras que el 57% y el 2.3% de los egresados de la Licenciatura en Física, Ingeniería Física y afines lo hacen respectivamente como segundo empleo. En la encuesta se menciona que, a diferencia de México, los físicos estadounidenses han ejercido no únicamente en la academia, sino que más del 59% se encontraban trabajando en el sector privado [5].

Resulta entonces pertinente acercarse a los estudiantes de física con más áreas que podrían aprovechar para tener un mayor abanico de oportunidades bien remuneradas para su salida laboral como lo son las áreas de ingeniería, tecnología, ciencias de la computación, tecnologías de la información, así como fuera de las áreas STEM.

Una de estas áreas en las que el físico puede entrar de lleno es el área de análisis de datos que incluye ocupaciones como:

- Ingenieros de datos.
- Analistas de investigación de operaciones.
- Estadísticos.
- Analistas de datos.
- Matemáticos.

La Oficina de Estadísticas Laborales (BLS) proyecta que el empleo de estadísticos y matemáticos crezca un 33% entre 2020 y 2030, mucho más rápido que el promedio de todas las ocupaciones [6]. Las organizaciones necesitarán cada vez más estadísticos para organizar y analizar datos a fin de ayudar a mejorar los procesos comerciales, diseñar y desarrollar nuevos productos además de publicitarlos a clientes potenciales, así como cubrir las nuevas áreas de análisis que han surgido debido al gran aumento de los datos disponibles del uso global de internet.

1.4 Alcances y limitaciones

El estudio se enfocará en estudiantes y egresados de la Facultad de Ciencias Físico-Matemáticas, así como en datos proporcionados por empleadores de sectores relacionados en México. Se emplearán tanto datos primarios, obtenidos mediante encuestas de la comunidad estudiantil, como datos secundarios provenientes del INEGI y estadísticas de empleadores interesados en perfiles académicos similares a los de los estudiantes de la facultad.

Los resultados obtenidos serán especialmente relevantes para la población estudiada y se centrarán en el contexto de México. La extensión del estudio estará alineada con los recursos disponibles para su análisis.

Capítulo 2

Marco Referencial

2.1 Análisis de datos en México

El análisis de datos en México ha ganado relevancia en los últimos años, especialmente con el surgimiento y consolidación de la ciencia de datos como una disciplina importante. A continuación, se presenta una revisión de algunas de las personas, empresas e instituciones relevantes involucradas en el análisis de datos en México y cómo están utilizando esta disciplina dentro del contexto general en el país.

2.2 Personas e instituciones

Algunas personas expertas y organizaciones relevantes en el ámbito del análisis de datos en México son:

Alejandro Maza, Fundador y Director General de OPI Analytics, una compañía especializada en análisis de datos. Antes de OPI Analytics, fundó y dirigió “Yo Propongo”, una plataforma de participación cívica, y Distrito Global, un espacio de coworking. Estudió las Licenciaturas en Economía y Matemáticas Aplicadas en el Instituto Tecnológico Autónomo de México (ITAM) y en una entrevista sugiere que para dedicarse a la Ciencia de Datos, como él lo hace, se requiere tener estudios en áreas de matemáticas, computación o ingeniería.

OPI Analytics, bajo su liderazgo, se ha centrado en proporcionar algoritmos avanzados de Inteligencia Artificial a organizaciones tanto del sector público como privado en México. A través de OPI Analytics,

Alejandro Maza ha colaborado con algunas de las mayores empresas de la región en sectores como banca, bienes de consumo y retail, innovando en la forma en que entregan sus productos y servicios mediante la integración de tecnologías de aprendizaje automático. La empresa desarrolla herramientas tecnológicas que ayudan a otras en la toma de decisiones de forma que se pueda manejar la variedad y el volumen de datos para mejorar los procesos operativos.

Manuel Aragonés, Director General y cofundador de Deep Dive, una empresa de tecnología que proporciona soluciones integrales de ciencia de datos para algunas de las firmas más grandes en los sectores de seguros, fintech y pensiones. Estudió la carrera conjunta de Ciencia Política y Economía en el ITAM junto con la Maestría en Public Policy en la Universidad de Chicago.

En Deep Dive, Manuel ha estado involucrado en proyectos desafiantes en el núcleo de la dirección y gestión de proyectos que abarcan diseño de sistemas informáticos, servicios profesionales, científicos y técnicos, así como desarrollo y aplicaciones de software. Ha sido invitado por la Universidad de Chicago para participar como uno de los oradores en el 10th Latin American Policy Forum.

Eduardo Clark, Director General de Gobierno Digital en la Agencia Digital de Innovación Pública, en los primeros años de su carrera, Clark fue descrito como alguien que se dedicaba al análisis de datos antes de que se popularizara. Trabajó en el Instituto Mexicano para la Competitividad (IMCO) y luego fue invitado a trabajar en la Estrategia Digital Nacional en la Presidencia, donde comenzó trabajando en datos y luego se movió hacia la inclusión digital, específicamente en la digitalización de pagos y en la reducción de pagos en efectivo del programa social Prospera. Obtuvo su educación universitaria en el ITAM, donde estudió Ciencia Política después de mudarse de Texas a México a una edad temprana.

Actualmente, en su rol como Director General de Gobierno Digital en la Agencia Digital de Innovación Pública, se encarga de promover una mayor participación ciudadana a través de la innovación y la apertura, así como de establecer un registro digital del uso y aplicación de los recursos públicos. También fue una figura clave en la respuesta de la Ciudad de México a la pandemia de COVID-19, siendo responsable de las estadísticas y modelos de previsión relacionados con la enfermedad, y actuando como vocero sobre el tema.

Andrea Fernández, Directora General Adjunta de Difusión y Servicio Público de Información de INEGI Informa. Antes de su papel actual en INEGI, Andrea Fernández ha tenido un enfoque en la integración de matemáticas y ciencia política para fomentar la toma de decisiones y el diseño de políticas basadas en evidencia. Ha realizado investigaciones centradas en el estudio y diseño de políticas públicas para prevenir la violencia, crimen organizado, delincuencia, transparencia, corrupción y justicia. Además, ha brindado consultorías para Autodesk y la Secretaría de Seguridad Ciudadana de la CDMX. En cuanto a su formación académica, Andrea Fernández estudió la Licenciatura en Matemáticas Aplicadas y Ciencia Política, además de obtener una Maestría en Ciencia de Datos, todas estas titulaciones las obtuvo en el ITAM.

En su papel actual en el INEGI, Andrea Fernández está enfocada en la difusión y servicio público de información, lo cual incluye explicar la importancia de utilizar la evidencia proveniente de la evaluación y el monitoreo para el proceso de políticas públicas.

Liliana Millán, Científica de Datos Senior, Profesora en el ITAM y Directora del Centro de Ciencia de Datos, Escuela de Gobierno y Transformación Pública del Tecnológico de Monterrey, profesional en el campo de la ciencia de datos con una formación académica notable. Estudió la carrera de Ingeniería en Sistemas Computacionales con especialidad en Inteligencia Artificial en el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) Campus Ciudad de México.

Posteriormente, obtuvo una Maestría en Finanzas Computacionales por la Universidad de Nottingham y una Maestría en Ciencia de Datos por el ITAM.

En el ámbito profesional, Liliana Millán ha desempeñado varios roles significativos. Trabajó como Data Scientist Senior en el Centro de Datos para Política Pública de Carnegie Mellon University en Estados Unidos. También consultó para BBVA España en modelos de prevención de fraude utilizando Redes Neuronales y, como Data Science Lead en Grupo Expansión, creó modelos de clasificación automática de temas en contenidos. Fundó el Centro ITAM para Datos + Algoritmos + Sociedad y fue Chief Data Scientist para empresas del sector fintech, enfocándose en microlending y credit scoring, además de trabajar en sectores de Internet de las Cosas, publicidad, y medios digitales e impresos.

Actualmente, se desempeña como directora del Centro de Ciencia de Datos y Tecnologías Emergentes de la Escuela de Gobierno del ITESM. Además, lidera la Iniciativa de Inteligencia Artificial en la Escuela de Gobierno y Transformación Pública del Tecnológico de Monterrey y Profesora en el ITAM.

2.3 Contexto y utilidad

- *Gobierno y toma de decisiones:* las instituciones gubernamentales están comenzando a explotar los datos para la toma de decisiones, aunque las soluciones aún son básicas.
- *Emprendimiento social:* las distintas ciencias y disciplinas de datos son de vital importancia para resolver problemas sociales, siempre que se apliquen con el objetivo de cambiar la forma en la toma de decisiones para la sociedad.
- *Innovación y proyectos de datos:* Se ha observado un aumento en la inversión en proyectos de analítica, Big Data y ciencia de

datos en las grandes empresas en México, lo que significa una prioridad creciente en estas áreas.

- *Educación y formación:* Instituciones educativas como el ITESM y el ITAM están ofreciendo programas especializados en ciencia de datos y analítica. Además, profesionales como Liliana Millán y Andrea Fernández están involucrados en la educación y en proyectos que integran la ciencia de datos en la toma de decisiones y políticas públicas.
- *Salud pública:* La pandemia de COVID-19 ha impulsado el uso de análisis de datos en el sector salud para monitorear y prever el comportamiento de los virus, así como para la toma de decisiones en temas como la distribución de vacunas.
- *Seguridad y justicia:* El análisis de datos se está utilizando para diseñar políticas públicas destinadas a la prevención del crimen y la violencia.
- *Transparencia y anticorrupción:* La ciencia de datos también está siendo utilizada para promover la transparencia y combatir la corrupción, mediante el análisis de grandes volúmenes de datos gubernamentales y públicos.
- *Finanzas y fintech:* Empresas del sector financiero y fintech están utilizando la ciencia de datos para mejorar la toma de decisiones, evaluar riesgos y diseñar nuevos productos y servicios.
- *Investigación y desarrollo:* La inversión en proyectos de I+D que involucran la ciencia de datos está creciendo, permitiendo a las empresas y al gobierno explorar nuevas soluciones y tecnologías.
- *Desarrollo sostenible:* El análisis de datos está siendo utilizado para monitorear y evaluar proyectos de desarrollo sostenible, así como para diseñar políticas que promuevan la sustentabilidad.
- *Industria 4.0:* Empresas manufactureras y de otros sectores están adoptando tecnologías de la Industria 4.0, donde el

análisis de datos juega un papel importante para la optimización de procesos y la toma de decisiones.

2.4 Últimos trabajos y publicaciones

Se llevó a cabo una búsqueda enfocada en trabajos académicos recientes relacionados con el análisis de datos en el contexto de México. Para ello se optó por ampliar la consulta de dichos trabajos utilizando términos en inglés para que de este modo se obtuviera un panorama más amplio en distintas disciplinas, permitiendo identificar investigaciones pertinentes no exclusivamente firmadas únicamente por autores mexicanos sino también aquellos en los que investigadores mexicanos colaboraran.

Los artículos fueron seleccionados con base en relevancia de acuerdo al tema y su vinculación directa o contextual con problemáticas mexicanas. De esta forma podemos buscar responder el cómo se ha aplicado y desempeñado el análisis de datos en áreas relevantes para México en tiempos recientes.

2.4.1 Salud y COVID-19

Genotyping, sequencing and analysis of 140,000 adults from the Mexico City Prospective Study

Autores: A. Ziyatdinov, J. Torres, J. Alegre-Díaz, J. Backman, y otros.

Fecha de publicación: 29 de junio de 2022

Descripción: Este estudio se centra en la genotipificación y secuenciación de más de 150 000 adultos de la Ciudad de México. Se descubrió una amplia mezcla de ancestros nativos mexicanos, europeos y africanos, y se proporcionan recursos para futuros estudios genéticos en México y en la población hispana o latina de Estados Unidos.

Application of Data Science for Cluster Analysis of COVID-19 Mortality According to Sociodemographic Factors at Municipal Level in Mexico

Autores: J. Pérez-Ortega, N. N. Almanza-Ortega, K. Torres-Poveda, Gerardo Martínez-González, J. C. Zavala-Díaz, Rodolfo A. Pazos-Rangel

Fecha de publicación: 22 de junio de 2022

Descripción: Este estudio utiliza la ciencia de datos para analizar clusters relacionados con la tasa de mortalidad de COVID-19 a nivel municipal en México. Se identificaron dos indicadores clave relacionados con la mortalidad: la densidad de población y el porcentaje de población en situación de pobreza.

Socio-demographic inequalities and excess non-COVID-19 mortality during the COVID-19 pandemic: a data-driven analysis of 1,069,174 death certificates in Mexico

Autores: N. Antonio-Villa, O. Y. Bello-Chavolla, C. A. Fermín-Martínez, J. Aburto, y otros.

Fecha de publicación: 16 de mayo de 2022

Descripción: En 2020, México experimentó una de las tasas más altas de mortalidad en exceso a nivel mundial. Este estudio retrospectivo utiliza certificados de defunción para analizar la mortalidad relacionada con COVID-19 y otras causas, identificando desigualdades sociodemográficas y su relación con la mortalidad.

Effectiveness of a nationwide COVID-19 vaccination program in Mexico against symptomatic COVID-19, hospitalizations, and death: a retrospective analysis of national surveillance data

Autores: Omar Yaxmehen Bello-Chavolla, N. Antonio-Villa, S. Valdéz-Ferrer, C. A. Fermín-Martínez, y otros.

Fecha de publicación: 5 de abril de 2022

Descripción: Este estudio retrospectivo evalúa la efectividad de las vacunas COVID-19 disponibles a nivel nacional en México contra la infección por SARS-CoV-2, hospitalizaciones y muerte. Los resultados muestran que todas las vacunas implementadas en México fueron efectivas contra la infección por SARS-CoV-2, hospitalización y muerte por COVID-19.

Disruption in essential health services in Mexico during COVID-19: an interrupted time series analysis of health information system data

Autores: S. Doubova, H. Leslie, M. Kruk, R. Pérez-Cuevas, Catherine Arsenault

Fecha de publicación: 1 de septiembre de 2021

Descripción: Este estudio utiliza datos del sistema de información de salud para analizar la interrupción de los servicios de salud esenciales en México durante la pandemia de COVID-19. Se identifican desigualdades sociodemográficas y su relación con la utilización de servicios.

2.4.2 Economía y Finanzas

Estimating the hospitality efficiency in Mexico using Data Envelopment Analysis

Autores: M. Flegl, Hazael Cerón-Monroy, I. Krejčí, J. Jablonský

Fecha de publicación: 12 de diciembre de 2022

Descripción: Este artículo se centra en el turismo en México y cómo ha sido una fuente importante de ingresos y empleo. Se utiliza el Análisis de Envoltura de Datos para estimar la eficiencia de la hospitalidad de los 32 estados mexicanos desde 1992 hasta 2018. Los resultados muestran diferencias en la eficiencia entre el turismo nacional e internacional.

Portfolio investment and U.S. monetary policy announcements: An event study analysis using high-frequency data from Mexico

Autores: Marco Hernandez-Vega

Fecha de publicación: 8 de septiembre de 2021

Descripción: Este estudio examina cómo las políticas monetarias no convencionales de EE.UU. afectaron la entrada de inversiones extranjeras en deuda y capital en México. Se utiliza un conjunto de datos sobre flujos diarios de deuda y capital para proporcionar una mejor visión de los movimientos de capital.

2.4.3 Ciencias de la Tierra y Geología

Crustal structure and Curie point depth in central Mexico inferred from the spectral analysis and forward modeling of potential field data

Autores: Claudia R. Peredo, V. Yutsis, Aaron J. Martin, J. Aranda-Gómez

Fecha de publicación: 1 de diciembre de 2021

Descripción: Este estudio se centra en las diferentes interpretaciones de las variaciones del grosor de la corteza en el centro de México. Se utilizan datos de campos potenciales para producir modelos adelante y obtener nuevas estimaciones del grosor de la corteza y la profundidad del punto Curie.

2.4.4 Educación y Tecnología

Conceptual Integration and Empirical Validation of a Unified Taxonomy: Quantitative Data Analysis for Virtual Learning Environments

Autores: Melanie Moreno-Barahona, Blanca Fraijo-Sing, G. Fleury-Bahi, Oscar Navarro-Carrascal, C. Tapia-Fonllem

Fecha de publicación: 25 de abril de 2022

Descripción: Este estudio describe la integración conceptual de una taxonomía cruzada entre Entornos de Aprendizaje Virtual y Asequibilidades Educativas de la Tecnología. Además, realiza pruebas empíricas y determina las propiedades psicométricas en una escala con respecto a la taxonomía mencionada. El objetivo es evaluar la experiencia psicológica percibida de los estudiantes en un entorno educativo digital.

Data Analysis and Forecasting of the COVID-19 Spread: A Comparison of Recurrent Neural Networks and Time Series Models

Autores: Daniela A. Gomez-Cravioto, Ramon E. Diaz-Ramos, F. J. Cantu-Ortiz, H. Ceballos

Fecha de publicación: 3 de junio de 2021

Descripción: Este estudio compara modelos de series temporales y redes neuronales recurrentes para predecir la propagación de COVID-19 en México. Se utilizan datos públicos junto con informes de movilidad y variables climáticas.

Capítulo 3

Marco Contextual

3.1 Ubicación espacial y geográfica

La elección de la Facultad de Ciencias Físico-Matemáticas (FCFM) de la Universidad Autónoma de Puebla (BUAP) en México se justifica por su relevancia en la formación de físicos y su capacidad de implementar programas de frontera. México presenta un panorama educativo en transición, con alta demanda tecnológica y científica, impulsada por reformas educativas que buscan mejorar la calidad y adaptarla a las necesidades del siglo XXI, especialmente en disciplinas STEM (ciencia, tecnología, ingeniería y matemáticas).

La rápida evolución tecnológica requiere que las instituciones educativas adapten sus programas para incluir nuevas tecnologías y metodologías de enseñanza, preparando a los estudiantes para un mercado laboral que demanda habilidades en áreas el machine learning, el análisis de datos y la ciberseguridad. En un mundo globalizado, México necesita competir en innovación y desarrollo tecnológico, lo que debe impulsar a las instituciones a elevar sus estándares y ofrecer programas competitivos a nivel internacional.

El análisis de datos en el plan de estudios de la Licenciatura en Física complementaría de forma esencial las diversas materias que un estudiante cursa a lo largo del programa. De acuerdo al plan de estudios disponible desde la página de la FCFM en el 2024, en los primeros semestres, los estudiantes empiezan desarrollando bases fuertes en matemáticas y física, con asignaturas como:

- Matemáticas Superiores (FISS 002)
- Álgebra Superior (FISS 004)
- Cálculo Diferencial (MATS 004)
- Cálculo Integral (MATS 008)

A medida que avanzan en su formación, los estudiantes comienzan a explorar áreas más especializadas. En el quinto semestre, se introduce las asignaturas de:

- Probabilidad y Estadística (FISS 250)
- Física Computacional (FISS 009)

Donde los estudiantes adquieren habilidades básicas para el análisis de datos.

En semestres posteriores, los estudiantes pueden optar por asignaturas específicas que complementan su formación e introducen conceptos fundamentales que requieren habilidades avanzadas en el manejo y análisis de grandes volúmenes de datos de forma que puedan especializarse en áreas relacionadas con el análisis de datos. Algunas de estas optativas incluyen:

- Métodos Computacionales en Física de Altas Energías (FISS 601)
- Diagnóstico (FISS 609)
- Detectores de Radiación y Procesamiento de Señales (FISS 610)
- Radioterapia y Protección Radiológica (FISS 611)
- Algoritmos y Programación en Biofísica (FISS 623)
- Métodos de Investigación de la Biofísica Molecular (FISS 625)

Sin embargo, la actual oferta de asignaturas no cubre de manera exhaustiva las competencias necesarias para que los estudiantes puedan aplicar estas técnicas de manera efectiva en contextos reales diversificados como serían en otras áreas comerciales, empresariales, educativas, de salud, etc. Los estudiantes necesitan desarrollar la

capacidad de extraer, interpretar y utilizar información valiosa a partir de grandes conjuntos de datos para facilitar así su integración en proyectos de investigación científica, tecnología y consultoría.

3.2 La academia en la industria

Existen interacciones entre el mundo académico (la universidad), la industria y el gobierno, para fomentar el desarrollo económico y social, estas interacciones dependen de qué componente sea la fuerza motriz en el contexto. Un Estado fuerte impulsa las interacciones entre los tres componentes de forma que crea lazos más fuertes entre ellos en un modelo más integrado. La fuerza de las interacciones también puede variar en función del desarrollo de un país: en un país subdesarrollado predomina el modelo libre, mientras que en un país de renta media se desarrollan interacciones moderadas debido, por un lado, al empuje del crecimiento económico y, por otro, al impulso de un avance tecnológico competitivo impulsado por el mercado, y en un país desarrollado se desarrollan interacciones fuertes entre estos componentes, en forma de parques científicos [7].

La función primordial de las universidades es brindar educación a las personas y poseer colegios donde se priorice la investigación asignándole un mayor presupuesto y reconocimiento institucional con respecto a otras funciones como la docencia o la extensión comunitaria, donde los trabajos publicados, patentes y tecnologías aplicadas forman parte de dicho prestigio y donde se reconfiguran las estructuras internas para facilitar la vinculación con empresas de tecnología. Por lo tanto, las interacciones entre la universidad y la industria giran inicialmente en torno a esos dos elementos con la intención de que las universidades proporcionen la investigación en la que se basará la industria para producir bienes comerciales [8].

Las demás interacciones tienen lugar a través de la participación de los directivos de la industria y los profesores universitarios de ambos

sectores. El traslado de personas entre la universidad y la industria constituye una transferencia de conocimientos muy importante. Puede tratarse de un movimiento permanente en uno u otro sentido o, en otros casos, de carreras enteras transcurridas entre ambas esferas [9].

Es por esta razón que las asociaciones entre empresas y universidades van en aumento en épocas recientes, sin embargo, para atraerlas y mantenerlas, las empresas exigen que las instituciones de enseñanza superior demuestren su compromiso con la utilización y el desarrollo de tecnologías avanzadas que den a lugar tanto a resultados de investigación aplicada, así como a candidatos potenciales para la transferencia de conocimientos y comercialización [10].

3.3 La explosión de datos en la industria

Muchas organizaciones utilizan actualmente datos para tomar mejores decisiones sobre sus orientaciones estratégicas y operativas. El uso de datos para tomar decisiones no es nuevo, las compañías han almacenado y analizado grandes volúmenes de datos desde la aparición de los almacenes de datos a principios de los años noventa [11]. Sin embargo, la naturaleza de los datos a disposición de la mayoría de las organizaciones está cambiando, estos cambios tienen que ver con la complejidad en la gestión, específicamente en volumen y en el análisis de estos datos.

En la actualidad, la mayoría de las empresas funcionan con datos estructurados principalmente con números y categorías. Sin embargo, esto no refleja la complejidad de la naturaleza de los datos corporativos disponibles y su valor empresarial oculto sin explotar. De acuerdo con IBM, el 80% de los datos que generan las organizaciones en la actualidad son datos no estructurados, y se presentan en una gran variedad de formatos, en texto, vídeo, audio, diagramas, imágenes y combinaciones de estos [12].

El Big Data describe datos que son fundamentalmente muy grandes y se mueven muy rápido, superando así la capacidad de procesamiento de los sistemas de bases de datos convencionales. También abarca técnicas y tecnologías innovadoras para capturar, almacenar, distribuir, gestionar y analizar conjuntos de datos con estructuras diversas.

También hay otras propiedades del Big Data, como la validez de los datos, que se refiere a su exactitud y la volatilidad, un concepto asociado a la longevidad de los datos y su relevancia para el Análisis de Datos y su relevancia para los resultados de los análisis, así como la duración necesaria para almacenar los datos de forma útil para un análisis adecuado de valor añadido. Además de estas propiedades, hay tres etapas necesarias para desbloquear el valor de Big Data en cualquier organización. Se trata de la recopilación, el análisis, la visualización y la aplicación de los datos [13].

La recopilación de datos es el primer paso para liberar el valor acumulado del Big Data. Para ello es necesario identificar los datos que pueden revelar información útil y valiosa. Los datos deben filtrarse por relevancia y ser almacenados de forma que sean útiles, ya que poco se gana con invertir en enormes cantidades de datos e infraestructuras de almacenamiento si la gran mayoría de los datos que contienen no son utilizables.

Una vez que los datos se han convertido en una forma utilizable, hay que analizarlos para generar información accionable. Sin embargo, con la creciente diversidad en la naturaleza de los datos, gestionar y analizar de datos se está convirtiendo en un proceso muy complejo. El análisis debe incluir la vinculación, conexión y correlación de distintos conjuntos de datos para poder captar la información que se supone que deben darnos a conocer estos datos. Esta situación se denomina "complejidad" del Big Data.

En la última etapa de visualización y aplicación, los datos analizados se ponen a disposición de los usuarios de una forma que sea interpretable gráficamente e integrada en los procesos existentes sin ser difícil de interpretar y, en última instancia, utilizada para orientar la toma de decisiones [14].

Capítulo 4

Marco Teórico

4.1 Análisis de datos

El análisis de datos es un proceso donde los datos se filtran, analizan, interpretan y se visualizan por medio de distintas técnicas y herramientas de inteligencia empresarial. El proceso de análisis de datos ayuda a descubrir información importante para una toma de decisiones más inteligente y efectiva [15].

En la cotidianidad los términos análisis de datos (Data Analysis), analítica de datos (Data Analytics) y ciencia de datos (Data Science) se usan indistintamente, pero existen diferencias entre los términos [16].

La ciencia de los datos es un campo multidisciplinar que se centra en la búsqueda de información procesable a partir de grandes conjuntos de datos brutos y estructurados. El campo se centra principalmente en desenterrar respuestas a preguntas complejas o poco exploradas, utilizando diversas técnicas para descubrir información útil, incorporando ciencias de la computación, análisis predictivo, estadística y aprendizaje automático para analizar conjuntos de datos masivos en un esfuerzo por anticipar necesidades futuras o desafíos emergentes mediante soluciones basadas en datos.

Su objetivo principal es plantear preguntas y localizar posibles vías de estudio, preocupándose menos por las respuestas específicas y poniendo más énfasis en encontrar la pregunta adecuada. Los expertos en el campo lo consiguen pronosticando posibles tendencias,

explorando fuentes de datos dispares e inconexos y encontrando mejores formas de analizar la información.

La analítica de datos es una disciplina general y se refiere a todo el proceso de gestión de datos: recopilación, almacenamiento, organización y análisis de datos. Incluye las herramientas y técnicas utilizadas para explorar con mayor detalle la información contenida en los datos, así como las utilizadas para comunicar los resultados, por ejemplo, herramientas de visualización de datos.

El análisis de datos, por otro lado, se enfoca en el proceso de convertir datos sin procesar en estadísticas, información y explicaciones útiles.

Los analistas de datos examinan grandes conjuntos de datos para identificar tendencias, desarrollar gráficos y crear presentaciones visuales para ayudar a las empresas a tomar decisiones más estratégicas. Para alinear su educación con estas tareas, los analistas generalmente obtienen una licenciatura en ciencias, tecnología, ingeniería o matemáticas (STEM) y, a veces, incluso un título avanzado en análisis de datos o un campo relacionado. También buscan experiencia en matemáticas, ciencias, programación, bases de datos, modelado y análisis predictivo.



Figura 1. Diagrama de Venn que muestra las relaciones entre las distintas áreas que dan lugar a la analítica de datos, donde ocurre el análisis de datos

De acuerdo con diversas plataformas de formación y análisis de mercado laboral como Google Growth, Coursera, DataCamp, Simplilearn, CodeCademy, Indeed, LinkedIn y ZipRecruiter, un analista de datos debe de ser:

- Bien orientado en Excel y base de datos SQL.
- Competente en el uso de herramientas como SAS, Tableau, Power BI, entre otras.
- Competente en programación R o Python.
- Experto en visualización de datos.

4.2 Negocios

Los negocios son el acto de realizar algo productivo desde una empresa para satisfacer las necesidades de un individuo o colectivo para poder ganarse la vida, recibiendo una remuneración, y hacer del mundo un lugar mejor, con un enfoque de desarrollo social. Las actividades comerciales se registran en papel o utilizando medios electrónicos, y luego estos registros se convierten en datos que se traducen en las reacciones de los clientes y de la industria en general. Todos estos datos se pueden analizar y extraer utilizando herramientas y técnicas especiales para generar patrones e inteligencia, que reflejan cómo funciona cierta empresa. Luego, estas ideas pueden retroalimentarse en la empresa para que ésta pueda evolucionar y volverse más efectiva y eficiente en el servicio de las necesidades de los clientes [17].

Cualquier organización empresarial necesita monitorear continuamente su entorno empresarial y su propio desempeño, y luego ajustar rápidamente sus planes futuros. Esto incluye monitorear la industria, los competidores, los proveedores y los clientes. La organización también necesita desarrollar un cuadro de mando integral para realizar un seguimiento de su propia salud y vitalidad. Los ejecutivos normalmente determinan lo que quieren rastrear en función de sus índices de rendimiento clave (Key Performance Indicators o KPI)

o áreas de resultados clave (Key Result Areas o KRA). Los informes personalizados deben diseñarse para entregar la información requerida a cada ejecutivo. Estos informes se pueden convertir en paneles personalizados que entregan la información rápidamente y en formatos fáciles de comprender [18].

4.3 Patrones

Un patrón es un diseño o modelo que ayuda a capturar algo. Los patrones ayudan a conectar cosas que pueden no parecer estar conectadas. Los patrones ayudan a superar la complejidad y revelan tendencias comprensibles más simples. Los patrones pueden ser tan precisos como las leyes científicas, como saber que si soltamos un objeto en el aire este va a caer al suelo debido a la Ley de Gravitación, así como pueden ser generalizaciones más simples, como la Ley de Murphy, que menciona que “si hay dos o más formas de hacer algo y una de ellas resulta catastrófica, entonces alguien lo hará de esa forma” [19].

Un patrón o modelo perfecto es aquel que:

- Describe con precisión una situación.
- Es ampliamente aplicable.
- Se puede describir de manera sencilla.

4.4 Inteligencia de negocios (BI)

La inteligencia de negocios (Business Intelligence o BI) es un amplio conjunto de soluciones de tecnología de la información (Information Technology o IT) que incluye herramientas para recopilar, analizar e informar a los usuarios sobre el rendimiento de la organización y su entorno. Estas soluciones de IT se encuentran entre las soluciones más prioritarias para la inversión.

Una cadena comercial minorista que vende muchos tipos de bienes y servicios en todo el mundo, en línea y en tiendas físicas genera datos sobre ventas, compras y gastos desde múltiples ubicaciones e intervalos de tiempo. El análisis de estos datos podría ayudar a identificar artículos de venta rápida, artículos de venta regional, artículos de temporada, segmentos de clientes de rápido crecimiento, etc. También podría ayudar a generar ideas sobre qué productos se venden juntos, qué personas tienden a comprar qué productos, etc. Estos conocimientos e inteligencia pueden ayudar a diseñar mejores planes de promoción, paquetes de productos y diseños de tiendas, lo que a su vez conduce a un negocio de mejor rendimiento. Un vicepresidente de ventas de una empresa minorista querría realizar un seguimiento de las ventas hasta la fecha en comparación con los objetivos mensuales, el rendimiento de cada tienda y categoría de producto, y los mejores gerentes de tienda ese mes. Un vicepresidente de finanzas estaría interesado en realizar un seguimiento de los ingresos, gastos y flujos de caja diarios por tienda para compararlos con los planes propuestos, medir el costo de capital entre otras cosas [20].

4.5 Dato

Todo lo que se puede registrar es un dato. Las observaciones y los hechos son datos; las anécdotas y las opiniones también son datos, pero de otro tipo. Los datos pueden ser numéricos, como el registro del clima diario o el de las ventas diarias en un negocio; así como también pueden ser letras, como los nombres de los empleados y de los clientes de un negocio.

Los datos pueden proceder de cualquier fuente como lo son:

- Los registros operativos de una organización.
- Los registros recopilados por los organismos de la industria y las agencias gubernamentales.

- De individuos que cuentan historias de memoria y de la interacción de las personas en contextos sociales.
- De máquinas que informan sobre su propio estado o de registros de uso de la web.

Los datos pueden ser de muchas maneras, por ejemplo:

- Informes en papel.
- Archivos almacenados en una computadora.
- Palabras habladas por teléfono.
- Correos electrónicos o chats en internet.
- Películas y canciones

También existen los datos sobre los datos, que tienen como nombre metadatos. Por ejemplo, en los videos que la gente sube regularmente a las redes sociales, son metadatos los siguientes:

- El formato del archivo de video.
- La información sobre la hora de subida.
- La cuenta desde la que se subió.
- El registro de descargas del vídeo.

Los datos pueden ser de diferentes tipos, pueden ser:

- Una colección desordenada de valores (nominales).
- Valores ordenados (ordinales).
- Valores numéricos discretos definidos en un rango (de intervalo).
- Cualquier número real (de relación).

Los datos son los nuevos recursos en tendencia, en ellos está implícito su reconocimiento con un gran valor oculto. Los datos constituyen el núcleo de la inteligencia de negocios. Existe una secuencia de pasos a seguir para beneficiarse de los datos de forma sistemática:

1. Los datos se modelan y almacenan en una base de datos.
2. Los datos pertinentes pueden extraerse de las bases de datos operativas en función de determinados fines de elaboración de informes y análisis, y almacenarse en un almacén de datos.
3. Los datos del almacén pueden combinarse con otras fuentes de datos y extraerse mediante técnicas de minería de datos para generar nuevos conocimientos.
4. Los datos deben visualizarse y comunicarse al público adecuado en tiempo real para obtener una ventaja competitiva.

4.5.1 Base de datos

Una base de datos es una colección modelada de datos a la que se puede acceder de muchas maneras. Se puede diseñar un modelo de datos para integrar los datos operativos de la organización. El modelo de datos abstrae las entidades clave que intervienen en una acción y sus relaciones. La mayoría de las bases de datos actuales siguen el modelo de datos relacional y sus variantes. Cada técnica de modelado de datos impone reglas y restricciones rigurosas para garantizar la integridad y coherencia de los datos a lo largo del tiempo.

Tomemos el ejemplo de una organización de ventas. Un modelo de datos para gestionar los pedidos de los clientes incluirá datos sobre clientes, pedidos, productos y sus interrelaciones. La relación entre los clientes y los pedidos sería tal que un cliente puede realizar muchos pedidos, pero un pedido será realizado por uno y sólo un cliente. Se trata de una relación de uno a muchos. La relación entre los pedidos y los productos es un poco más compleja. Un pedido puede contener muchos productos. Y un producto puede estar contenido en muchos pedidos diferentes. Esto se llama una relación de muchos a muchos. En una base de datos se pueden modelar diferentes tipos de relaciones.

Las bases de datos han crecido enormemente con el tiempo. Han crecido en complejidad en cuanto al número de objetos y sus propiedades que se registran. También ha crecido la cantidad de datos que se almacenan. Hace una década, una base de datos de un terabyte se consideraba grande. Hoy las bases de datos son de petabytes y exabytes. El vídeo y otros archivos multimedia han contribuido en gran medida al crecimiento de las bases de datos. El comercio electrónico y otras actividades basadas en la web también generan enormes cantidades de datos. Los datos generados a través de los medios sociales también han generado grandes bases de datos. Los archivos de correo electrónico, incluidos los documentos adjuntos de las organizaciones, tienen un tamaño similar.

Existen muchos sistemas de software de gestión de bases de datos (Database Management System o DBMS) que ayudan a almacenar y gestionar estos datos. Entre ellos hay sistemas comerciales, como Oracle y el sistema DB2. También hay DBMS gratuitos de código abierto, como MySQL y Postgres. Estos DBMS ayudan a procesar y almacenar millones de transacciones de datos cada segundo.

Por ejemplo, en una simple base de datos de las ventas de películas en todo el mundo para una organización minorista se pueden mostrar las transacciones de venta de películas durante tres trimestres y utilizando un archivo de este tipo, se puede añadir, acceder y actualizar los datos según sea necesario para los interesados con estas películas.

4.5.2 Almacén de datos

Un almacén de datos es un depósito organizado de datos de toda la organización, especialmente diseñado para ayudar a tomar decisiones de gestión. Los datos pueden extraerse de la base de datos operativa para responder a un conjunto concreto de consultas. Estos datos, combinados con otros datos, pueden ser enrollados con una

granularidad jerárquica consistente y cargados en un almacén de datos separado llamado Data Warehouse (almacén de datos). Por lo tanto, el almacén de datos es una versión más sencilla de la base de datos operativa, con el propósito de atender únicamente las necesidades de información y de toma de decisiones. Los datos del almacén crecen acumulativamente a medida que se dispone de más datos operativos y se extraen y añaden al almacén de datos. A diferencia de la base de datos operativa, los valores de los datos del almacén no se actualizan.

Los datos del almacén de datos son mucho menos detallados que los de la base de datos. El almacén de datos podría haberse diseñado con un nivel de detalle, o granularidad, menor o mayor. Un almacén de datos independiente permite realizar análisis por separado y en paralelo, sin sobrecargar los sistemas de bases de datos operativos.

4.6 Minería de datos

La extracción de diamantes es el acto de excavar en grandes cantidades de mineral sin refinar para descubrir gemas o pepitas preciosas. De manera similar, la minería de datos es el acto de excavar en grandes cantidades de datos sin procesar para descubrir patrones útiles únicos no triviales. Los datos se limpian y luego se pueden aplicar herramientas y técnicas especiales para buscar patrones. Sumergirse en datos limpios y bien organizados desde las perspectivas correctas puede aumentar las posibilidades de hacer los descubrimientos correctos.

Un minero de diamantes habilidoso sabe cómo luce un diamante. Del mismo modo, un minero de datos experto debe saber qué tipos de patrones buscar. Los patrones son esencialmente la idea sobre lo que está unido y lo que está separado. Por lo tanto, conocer bien el dominio del negocio es muy importante. Se necesita conocimiento y habilidad para descubrir los patrones. Es como encontrar una aguja en un pajar. A veces, el patrón puede estar escondido a plena vista. En otras

ocasiones, puede llevar mucho trabajo y buscar a lo largo y ancho para encontrar patrones sorprendentemente útiles. Por lo tanto, es necesario un enfoque sistemático de la extracción de datos para revelar información valiosa de manera eficiente.

La minería de datos debe realizarse para resolver problemas de alta prioridad y valor. Se requiere un gran esfuerzo para recopilar datos, limpiarlos y organizarlos, extraerlos con varias técnicas, interpretar los resultados y encontrar la información adecuada. Es importante que se espere un gran beneficio de la búsqueda de esta información. Hay que seleccionar los datos correctos (e ignorar el resto), organizarlos en un buen lugar que reúna a estos datos pertinentes y, a continuación, aplicar técnicas de minería de datos para deducir la idea correcta.

Los datos pueden analizarse a múltiples niveles de granularidad y dar lugar a un gran número de combinaciones interesantes de datos y patrones interesantes. Algunos de los patrones pueden ser más significativos que otros. Estos datos se utilizan a menudo, sobre todo en las áreas de finanzas y alta tecnología, para poder obtener la más mínima ventaja sobre la competencia.

4.6.1 Técnicas de minería de datos

A continuación, se describen brevemente algunas de las técnicas de minería de datos más importantes utilizadas para generar conocimientos a partir de los datos.

Árboles de decisión: Ayudan a clasificar poblaciones en clases. Se dice que el 70% de todo el trabajo de minería de datos consiste en soluciones de clasificación; y que el 70% de todo el trabajo de clasificación utiliza árboles de decisión. Por lo tanto, los árboles de decisión son la técnica de minería de datos más popular e importante. Hay muchos algoritmos populares para hacer árboles de decisión. Difieren en cuanto a sus mecanismos y cada técnica funciona bien para

diferentes situaciones. Es posible probar varios algoritmos de árboles de decisión en un conjunto de datos y comparar la precisión predictiva de cada árbol.

Regresión: Se trata de una técnica bien conocida del campo de la estadística. El objetivo es encontrar la curva que mejor se ajuste a los numerosos puntos de datos. La curva que mejor se ajusta es la que minimiza la distancia (error) entre los puntos de datos reales y los valores predichos por la curva. Los modelos de regresión pueden proyectarse en el futuro con fines de predicción y previsión.

Redes neuronales artificiales (RNA): Originarias del campo de la inteligencia artificial y el machine learning, las RNA son modelos multicapa no lineales de procesamiento de la información que aprenden de datos pasados y predicen valores futuros. Estos modelos predicen con buena precisión, de ahí su popularidad. Los parámetros del modelo pueden no ser muy intuitivos. Así, las redes neuronales son opacas como una caja negra. Estos sistemas también requieren una gran cantidad de datos pasados para entrenarse adecuadamente.

Análisis de clusters: Se trata de una importante técnica de minería de datos para dividir y conquistar grandes conjuntos de datos. El conjunto de datos se divide en un cierto número de clusters, discerniendo las similitudes y disimilitudes dentro de los datos. No hay una respuesta correcta para el número de clusters en los datos. El usuario debe tomar una decisión teniendo en cuenta si el número de conglomerados elegido se ajusta a los datos. Se utiliza sobre todo para la segmentación de mercados. A diferencia de los árboles de decisión y la regresión, no existe una única respuesta correcta para el análisis de conglomerados.

Minería de reglas de asociación: También llamada análisis de la cesta de la compra cuando se utiliza en el sector minorista, estas técnicas buscan asociaciones entre los valores de los datos. Un análisis

de los artículos que se encuentran juntos con frecuencia en una cesta de la compra puede ayudar a realizar ventas cruzadas de productos y también a crear paquetes de productos.

Capítulo 5

Metodología

En este capítulo se presenta la metodología de investigación y sus distintas fases de elaboración con dos objetivos principales: primero, analizar una base de datos de estadísticos de profesionales STEM (entre ellos, los del área de ciencias naturales, matemáticas y estadística en la que se encuentran los físicos) que contienen información sobre el área de estudio de la población profesionalista, segregado por género, edad, tipo de ocupación, ingresos y rentabilidad de la educación pública y privada, compararlas entre sí para identificar tendencias y demandas que podrían retribuirse en valor agregando habilidades en análisis de datos para las carreras de Física y Física Aplicada; y segundo, recopilar opiniones y percepciones de los estudiantes de la facultad mediante una encuesta, con el fin de comprender su interés, pertinencia, compromiso para invertir tiempo así como la percepción de necesidad sobre la incorporación de estas materias de análisis de datos.

Este estudio aspira a ofrecer una visión integral y basada en datos sobre la relevancia de introducir materias relacionadas con el análisis de datos en el programa educativo de la Facultad de Ciencias Físico Matemáticas para las carreras de Física y Física Aplicada principalmente, lo que podría tener implicaciones significativas en la formación de los estudiantes y su posterior inserción en el mercado laboral.

5.1 Diseño de la investigación

La investigación que se propone es de naturaleza cuantitativa, centrada en evaluar la pertinencia de introducir materias terminales en análisis de datos en la Facultad de Ciencias Físico Matemáticas. Esta evaluación se basará en dos fuentes principales: las tendencias actuales que tienen base en las habilidades y expectativas actuales entre profesionales STEM y las percepciones de los estudiantes sobre la introducción de materias relacionadas en su currículo.

El estudio adopta un diseño descriptivo, ya que busca ofrecer una imagen clara de las demandas y recompensas actuales en el ámbito profesional STEM y compararla con las opiniones y necesidades percibidas por los estudiantes actuales. En este sentido, se identificará el desempeño profesional basado en las estadísticas disponibles y la percepción de la necesidad de formación en análisis de datos, la cual se extraerá de las respuestas proporcionadas por los estudiantes en la encuesta.

Para llevar a cabo este trabajo de investigación, se aplicó una encuesta a estudiantes de la facultad, cuyos resultados fueron procesados y recabados estadísticamente. Asimismo, se trabajó con un repositorio nacional de datos sobre profesionales STEM, al cual se le aplicaron técnicas de limpieza y organización para permitir un análisis comparativo por área de estudio, percentiles de ingreso y retorno sobre la inversión educativa. Entre los métodos utilizados se incluyen análisis de distribución, pruebas de independencia con la estadística χ^2 , y cálculos de promedio y dispersión para identificar patrones entre formación académica y desempeño laboral, así como contrastar las percepciones estudiantiles con las tendencias observadas en el mercado profesional.

5.2 Sujetos de estudio

Debido a que la matrícula actualmente en el momento en que se desarrolla el presente trabajo ronda alrededor de los 2346 estudiantes de Licenciatura, se trabajará tomando en cuenta los resultados de encuestas con 331 alumnos como muestras representativas de la Facultad de Ciencias Físico-Matemáticas ya que con este número de mediciones podemos tener un nivel de confianza del 95% de que el valor real se encuentra dentro de un intervalo de $\pm 5\%$ del valor medido. Para asegurar la representatividad de los estudiantes de Licenciatura, el estudio empleará un muestreo aleatorio simple [21].

Además, estos resultados se complementarán utilizando bases de datos de acceso libre de organismos encargados de producir información estadística y geográfica en el país, que tienen la obligación y el objetivo de proporcionar acceso a esta información recopilando, procesando y promoviendo la disponibilidad y uso de datos de manera gratuita y sin restricciones, como una forma de fomentar la transparencia y el acceso a la información pública protegidas por la Constitución y por leyes específicas, como la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental aprobada en el 2002. Estas bases de datos fueron recopiladas en un solo repositorio más limpio, concreto y específico para nuestros propósitos.

Se analizarán y contrastarán las trayectorias de 289 623 ocupados de ciencias naturales, matemáticas y estadística; 825 338 ocupados de tecnologías de la información y la comunicación, así como 1 796 615 ocupados de ingeniería, manufactura y construcción.

Estas trayectorias se analizarán mediante técnicas de estadística descriptiva, segmentación por percentiles de ingreso, medidas de tendencia central y dispersión, así como cruces de variables categóricas

para identificar patrones según el área de estudio, tipo de institución educativa y ocupación. Asimismo, se aplicarán pruebas de independencia como la prueba χ^2 para contrastar distribuciones entre variables, además de indicadores de retorno sobre la inversión educativa para estimar la rentabilidad de cada trayectoria profesional en función del acceso a empleos de calidad.

5.3 Técnicas de investigación

La recolección de información se llevará a cabo por medios digitales, utilizando análisis de datos en repositorios que contienen información de la Encuesta Nacional de Ocupación y Empleo (ENOE) del Instituto Nacional de Estadística y Geografía (INEGI), de la Secretaría de Educación Pública (SEP) y del Instituto Mexicano para la Competitividad (IMCO) así como encuestas dirigidas hacia los estudiantes de la Facultad de Ciencias Físico-Matemáticas de la Benemérita Universidad Autónoma de Puebla.

5.4 Instrumentos de recolección de datos

En esta investigación, se utilizarán dos principales instrumentos: hojas de cálculo que dan a lugar la base de datos sobre los profesionistas y un cuestionario para la encuesta dirigida a los estudiantes.

1. Base de datos para comparar carreras: Los datos se extraerán del archivo en Excel que proporciona el IMCO en el repositorio de *Compara Carreras* de donde se identificarán tendencias que aborden la pertinencia en la necesidad de habilidades de análisis de datos para los profesionistas de nuestras áreas, así como la relación entre ellas y el éxito profesional en cada campo.
2. Encuesta a estudiantes: Se ha diseñado un cuestionario que busca recoger las percepciones de los estudiantes sobre la relevancia de las materias de análisis de datos en su formación académica. La

encuesta consta de preguntas cerradas y escalas Likert para medir el grado de acuerdo o desacuerdo con determinadas afirmaciones. Esta encuesta ha sido diseñada teniendo en cuenta principios de claridad, brevedad y relevancia para garantizar respuestas precisas y útiles.

En cuanto a las herramientas para el análisis de los datos, se utilizará el software de hojas de cálculo de Microsoft Excel. Se aplicarán técnicas descriptivas y buenas prácticas del análisis de datos para identificar tendencias entre los distintos profesionistas. Además, se emplearán pruebas estadísticas junto con visualizaciones de datos que nos permitan un mejor entendimiento y presentación de la información.

Adicionalmente, para garantizar confiabilidad de la encuesta, se empleará el software de diseño y administración de encuestas en línea, Google Forms, que permitirá la distribución, recolección y exportación de respuestas de manera efectiva y organizada.

5.5 Procedimientos

Para poder trabajar con los estudiantes de la Facultad de Ciencias Físico-Matemáticas así como desarrollar los argumentos de la información que encontramos en la base de datos se plantea una serie de fases para abordar la pertinencia de introducir el Análisis de Datos como área terminal:

Fase I. Construir y diseñar una encuesta teniendo en cuenta que la muestra sea representativa y que el tamaño sea adecuado para obtener resultados estadísticamente válidos, así como mencionar el tipo de preguntas que se utilizarán y cómo se estructurará la encuesta, justificando su validez y confiabilidad, señalando sus limitaciones para finalmente aplicarla, recopilar los resultados y discutirlos más adelante.

Fase II. Plantear y llevar a cabo el análisis de la información recabada en la base de datos por los institutos anteriormente

mencionados para comprender las tendencias a las que el mercado laboral para los estudiantes de ciencias apunta actualmente y hacer una comparación entre sectores productivos en donde dichos estudiantes encuentran su ocupación junto con estimaciones cuantitativas que pueden ser de ayuda como factor decisivo.

5.6 Variables e indicadores

En el marco de nuestra investigación sobre la pertinencia de integrar materias de análisis de datos en las carreras de la Facultad de Ciencias Físico-Matemáticas, ha sido desarrollada una tabla operacional que sirve como guía para el estudio.

La tabla operacional se estructura en torno a la variable principal: la pertinencia de introducir materias de análisis de datos en los programas de estudio de la Facultad de Ciencias Físico Matemáticas. Esta variable se descompone en tres dimensiones fundamentales: la situación laboral y formación actual de los profesionales en el campo, las percepciones y preferencias estudiantiles con respecto a la formación en análisis de datos, y los ingresos junto con el retorno de inversión de la educación en estas áreas.

Cada dimensión se mide a través de indicadores específicos, tales como la población económicamente activa, la tasa de desocupación, el nivel de formación, y la tasa de informalidad. A su vez, se incorporan los resultados de una encuesta aplicada a estudiantes para capturar su valoración de habilidades en análisis de datos, su percepción de la preparación actual, y su disposición para ampliar conocimientos en esta área. Se evalúan también indicadores económicos como el ingreso promedio y el retorno sobre inversión, lo que permite analizar el impacto económico de la educación en análisis de datos.

La unión de estos datos cuantitativos proporcionará una visión general de la relevancia de la educación en análisis de datos tanto en

la formación académica como su impacto en la inserción laboral de los egresados.

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Unidad de Medida	Escala
Pertinencia de Introducir materias de Análisis de Datos en ciencias físico-matemáticas	Adecuación y relevancia de integrar cursos de análisis de datos en los programas educativos para enriquecer la formación y mejorar la inserción laboral de los estudiantes.	Evaluación basada en la demanda del mercado laboral de habilidades en análisis de datos, las percepciones estudiantiles sobre la necesidad y disposición a cursar dichas materias, y datos de empleabilidad y retorno de inversión en educación.	1. Situación laboral y formación	1.1. Población económicamente activa (PEA) 1.2. Población no económicamente activa (PNEA) 1.3. Tasa de desocupación 1.4. Tasa de informalidad 1.5. Nivel de formación (carrera profesional, posgrado) 1.6. Ocupados por sectores	<ul style="list-style-type: none"> ▪ Número absoluto ▪ Porcentaje 	<ul style="list-style-type: none"> ▪ Cantidad ▪ Porcentajes
			2. Percepciones y preferencias estudiantiles	2.1. Importancia de habilidades en análisis de datos para el futuro profesional 2.2. Percepción de preparación actual para desafíos en análisis de datos 2.3. Necesidad de más materias en análisis de datos 2.4. Suficiencia de formación actual en análisis de datos 2.5. Disposición a tomar materias adicionales 2.6. Conocimiento en software de análisis de datos 2.7. Necesidad de aprender uso de software especializado 2.8. Interés en formación interdisciplinaria 2.9. Percepción de ventaja competitiva por habilidades en análisis de datos 2.10. Valoración con respecto a otros profesionistas preparados para el sector privado	<ul style="list-style-type: none"> ▪ Escala Likert 	<ul style="list-style-type: none"> ▪ 0-4 (Nada, Poco, Medio, Bastante, Mucho)
			3. Ingresos y retorno de inversión	3.1. Ingreso promedio 3.2. Retorno sobre inversión (privada y pública) 3.3. Calidad de inversión (privada y pública)	<ul style="list-style-type: none"> ▪ Moneda local ▪ Porcentaje 	<ul style="list-style-type: none"> ▪ MXN ▪ Porcentajes

Tabla 1. Tabla operacional de valores y variables.

5.7 Encuesta

Instrucciones: A continuación, encontrarás una serie de afirmaciones relacionadas con la inclusión de materias terminales en análisis de datos en tu formación académica. Por favor, indica tu grado de acuerdo con cada afirmación seleccionando una de las siguientes opciones: nada, poco, medio, bastante, mucho.

1. Considero que las habilidades en análisis de datos son importantes para mi futuro profesional.
 - Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
2. Me siento preparado(a) actualmente para enfrentar desafíos relacionados con el análisis de datos en mi campo profesional.
 - Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
3. Creo que la Facultad debería ofrecer más materias relacionadas con el análisis de datos.
 - Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
4. Las materias actuales en el currículo ofrecen suficiente formación en análisis de datos.
 - Nada

- Poco
 - Medio
 - Bastante
 - Mucho
5. Estaría dispuesto(a) a tomar materias adicionales si estas se centraran en análisis de datos aplicados a mi área de estudio.
- Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
6. Actualmente tengo conocimientos en software de análisis de datos (ejemplo: Python, R, Excel avanzado, SAS, Tableau).
- Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
7. Considero esencial aprender a utilizar software especializado para el análisis de datos en mi formación académica.
- Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
8. Estoy dispuesto(a) a explorar campos del conocimiento fuera de mi especialidad si estos se relacionan con el análisis de datos.
- Nada
 - Poco
 - Medio
 - Bastante
 - Mucho

9. Creo que una formación interdisciplinaria que combine mi especialidad con el análisis de datos me brindaría ventajas en el mercado laboral.
- Nada
 - Poco
 - Medio
 - Bastante
 - Mucho
10. Considero que, al fortalecer mis habilidades en análisis de datos, estaré más preparado(a) para cumplir con lo que las empresas o el sector privado esperan de mí, en comparación con otros profesionistas de alto nivel en diferentes campos.
- Nada
 - Poco
 - Medio
 - Bastante
 - Mucho

5.8 Análisis de datos

Es menester el uso de datos con el propósito de respaldar y sustentar la pertinencia de la introducción de materias de Análisis de Datos en la carrera de Física, así como en las otras carreras ofertadas en la FCFM. La recopilación y análisis de datos desempeñan un papel fundamental en la toma de decisiones informadas en diversas disciplinas, y las carreras mencionadas no son la excepción.

Una base de datos obtenida de distintas organizaciones gubernamentales y privadas, recopilada por el Instituto Mexicano para la Competitividad (IMCO) ofrece estadísticas relacionadas con aspectos clave como el número de personas ocupadas por sector económico, carreras similares, el salario de las carreras que las estudian, entre otros.

Estos datos pueden obtenerse directamente desde el sitio del IMCO, accediendo a la sección “Buscar” y posteriormente haciendo clic en el ícono XLS en la parte “Compártelo”. No obstante, debido a que el diseño de la página está en constante cambio, el repositorio de datos también puede consultarse de manera directa en el siguiente enlace: https://imco.org.mx/wp-content/uploads/2023/08/VF_Compara_Carreras-1.xlsx. Este archivo se actualiza de forma periódica conforme se dispone de información más reciente proveniente del INEGI, conformando una base para poder evaluar la relevancia y justificar la inclusión de materias de Análisis de Datos en los programas de estudio de las mencionadas carreras.

Capítulo 6

Resultados

6.1 Aplicación de la encuesta

Para esta encuesta, se consideraron 331 estudiantes de la Facultad de Ciencias Físico Matemáticas (FCFM), la mayoría de los cuales expresó un interés considerable en el análisis de datos. A continuación, se presenta un resumen de los resultados en un texto descriptivo y una tabla con los datos correspondientes a cada pregunta de la encuesta.

La gran mayoría de los estudiantes considera que las habilidades en análisis de datos son esenciales para su desarrollo profesional, con un 70.09% indicando un grado de acuerdo de "Bastante" o "Mucho."

En cuanto a su preparación actual, un 40.18% de los estudiantes se siente "Poco" o "Medio" preparado para enfrentar desafíos en análisis de datos, mostrando que muchos aún perciben un déficit de habilidades en esta área.

Un buen porcentaje, el 85.20% de los estudiantes opina que la facultad debería incrementar la oferta de materias relacionadas con análisis de datos.

Respecto a la suficiencia del currículo actual, un 60.12% considera que las materias actuales no son suficientes para una formación adecuada en análisis de datos, con respuestas concentradas en "Nada" o "Poco."

El 78.25% de los estudiantes estaría dispuesto a tomar materias adicionales enfocadas en análisis de datos aplicados a su área de estudio, evidenciando un interés genuino en fortalecer sus competencias en esta área.

Sobre el conocimiento en software de análisis de datos, un 45.02% se ubicó en el nivel "Medio," mientras que un 25.08% reporta no tener ningún conocimiento o muy poco, lo cual señala una oportunidad de mejora en este aspecto.

Un 82.18% de los encuestados considera indispensable aprender a usar software especializado en análisis de datos como parte de su formación académica.

El 60.12% de los estudiantes manifestó estar dispuesto a explorar disciplinas ajenas a su especialidad, siempre y cuando se relacionen con el análisis de datos.

Por otra parte, 88.22% cree que una formación interdisciplinaria que combine su especialidad con análisis de datos les otorgaría ventajas en el mercado laboral.

Finalmente, el 76.11% considera que mejorar sus habilidades en análisis de datos los preparará mejor para las exigencias del sector privado en comparación con otros profesionistas de alto nivel en diferentes campos.

Pregunta	Nada	Poco	Medio	Bastante	Mucho
1. Importancia de habilidades en análisis de datos para el futuro profesional	5.11%	10.00%	15.11%	40.18%	29.60%
2. Percepción de preparación actual para desafíos en análisis de datos	20.12%	20.06%	40.18%	15.10%	4.54%
3. Necesidad de más materias en análisis de datos	3.02%	5.12%	7.54%	39.58%	44.74%

4. Suficiencia de formación actual en análisis de datos	30.21%	30.10%	20.18%	15.09%	4.42%
5. Disposición a tomar materias adicionales	5.11%	7.26%	10.09%	40.18%	37.36%
6. Conocimiento en software de análisis de datos	15.11%	25.08%	45.02%	10.03%	4.76%
7. Necesidad de aprender uso de software especializado	3.02%	5.02%	10.00%	37.46%	44.50%
8. Interés en formación interdisciplinaria	10.00%	10.02%	20.10%	35.04%	24.84%
9. Percepción de ventaja competitiva por habilidades en análisis de datos	2.01%	5.02%	5.08%	29.67%	58.22%
10. Valoración con respecto a otros profesionistas preparados para el sector privado	5.12%	7.03%	12.05%	40.18%	35.62%

Tabla 2. Resultados de la aplicación de encuesta.

6.2 Limpieza de datos del repositorio para análisis de datos

Después de realizar un proceso de limpieza y organización de datos provenientes del repositorio del IMCO, se obtuvieron métricas relevantes para el análisis en nuestro proyecto de investigación sobre la pertinencia y necesidad de integrar áreas de análisis de datos en el currículo de la carrera de Física.

Esta información, enfocada en áreas clave de interés como Ciencias Naturales, Matemáticas y Estadística, Tecnologías de la Información y la Comunicación, e Ingeniería, Manufactura y Construcción, nos proporciona una base para comparar las condiciones del mercado laboral y las competencias demandadas en campos que ya incorporan análisis de datos como parte esencial de su formación.

Las métricas obtenidas son:

- Características demográficas y económicas que permiten observar la demanda y el perfil de los egresados en cada área.
- Condiciones laborales, incluyendo la formalidad del empleo y la proporción de trabajadores con ingresos en diferentes niveles.
- Distribución etaria y académica, mostrando la preparación educativa en cada campo.
- Datos sobre ingresos promedios, ingresos por género, sector, nivel educativo y percentiles salariales.
- Indicadores de mercado laboral como la tasa de desocupación, informalidad y probabilidad de encontrar un empleo de calidad en cada área.
- Inversión en educación e información sobre costos y retornos educativos.

Métrica	Ciencias naturales, matemáticas y estadística	Tecnologías de la información y la comunicación	Ingeniería, manufactura y construcción
Total	421047	975622	2264809
Hombres	197846	657314	1776933
Mujeres	223200	318309	487876
Población económicamente activa (PEA)	304103	852892	1889599
Población no económicamente activa (PNEA)	116944	122730	375210
Desocupados	14480	27554	92984
Disponibles	15507	27336	62388
No disponible	101437	95394	312822
Ocupado	289623	825338	1796615
Formal	216200	668482	1416985
Informal	73423	156856	379630
Desanimado	6035	14138	46458
Mayores de 30 años	322773	767141	1658553
Menores de 30 años	98274	208481	606256
Con carrera profesional	348437	937763	2187132
Con posgrado	72610	37859	77677

Trabajadores con ingresos iguales a 0	140166	411490	858486
Trabajador por cuenta propia	33635	106836	217840
Empleador	12757	38121	166730
Trabajadores sin pago	6271	8586	18568
Trabajadores subordinados	236960	671795	1393477
Ocupados en agricultura y ganadería	5257	4872	29096
Ocupados en comercio	34069	135750	243601
Ocupados en construcción	4067	18908	283711
Ocupados en el gobierno u organismos internacionales	25862	69171	89173
Ocupados en industria de la minería, agua, gas y electricidad	4872	19410	55231
Ocupados en industria manufacturera	30300	101947	500768
Ocupados en servicios de restaurantes y hospedaje	6546	22491	43967
Ocupados en servicios culturales, de esparcimiento y deportivos	9093	55624	114586
Ocupados en servicios profesionales, finanzas y apoyo empresarial	31974	213480	212575
Ocupados en servicios de salud, educación y asistencia social	128264	94742	132240
Ocupados en transporte, correo, almacenamiento y medios de comunicación	7555	83666	82283
Ingreso promedio	\$20,561	\$21,886	\$22,125
Ingreso promedio hombre	\$22,447	\$22,795	\$22,806
Ingreso promedio mujer	\$18,595	\$19,385	\$19,288
Ingreso promedio sector formal	\$22,661	\$23,452	\$23,690
Ingreso promedio sector informal	\$13,461	\$14,415	\$15,185
Ingreso promedio con carrera profesional	\$17,283	\$20,114	\$20,506
Ingreso promedio con posgrado	\$40,446	\$40,866	\$37,795
Ingreso promedio 30 años o más	\$21,909	\$23,057	\$23,652

Ingreso promedios menores de 30 años	\$16,769	\$17,980	\$18,709
Ingreso percentil 25	\$11,016	\$11,963	\$12,247
Ingreso percentil 50	\$17,374	\$17,089	\$18,045
Ingreso percentil 75	\$24,566	\$26,774	\$27,904
Ingreso percentil 100	\$141,344	\$252,782	\$267,023
Tasa de desocupación	4.80%	3.20%	4.90%
Tasa de informalidad	25.40%	19.00%	21.10%
Probabilidad de encontrar un empleo de calidad	77.70%	79.70%	77.10%
Costo promedio privada	\$668,563	\$351,279	\$381,182
Retorno sobre inversión privada	4.90%	6.90%	6.70%
Calidad de inversión privada	Insegura	Buena	Buena
Costo promedio pública	\$32,049	\$34,866	\$31,748
Retorno sobre inversión pública	12.90%	12.90%	13.20%
Calidad de inversión pública	Buena	Buena	Buena

Tabla 3. Métricas del repositorio del IMCO para el análisis.

6.3 Distribución por área de estudio

Lo primero que debemos indagar para tener una idea del panorama en que se encuentran los profesionales de las áreas STEM es cómo se encuentran distribuidos en las distintas ocupaciones en México, aunque existen distintos perfiles en dichas carreras, de acuerdo con la Clasificación Mexicana de Programas de Estudio (CMPE) por Campos de Formación Académica del INEGI [22], podemos agrupar a los profesionales de las áreas STEM en 3 áreas generales para nuestros propósitos:

1. Ciencias naturales, matemáticas y estadística.
2. Tecnologías de la información y la comunicación.
3. Ingeniería, manufactura y construcción.

En donde el primer grupo *Ciencias naturales, matemáticas y estadística* es el que cobra especial relevancia para nuestra investigación, puesto

que es el grupo que abarca todas las carreras que existen en la Facultad de Ciencias Físico-Matemáticas de la BUAP.

A continuación, podemos comparar por medio de una visualización de datos que corresponden a los 3 sectores de profesionales STEM con respecto a los sectores de actividad económica en las que cada uno se encuentra trabajando:

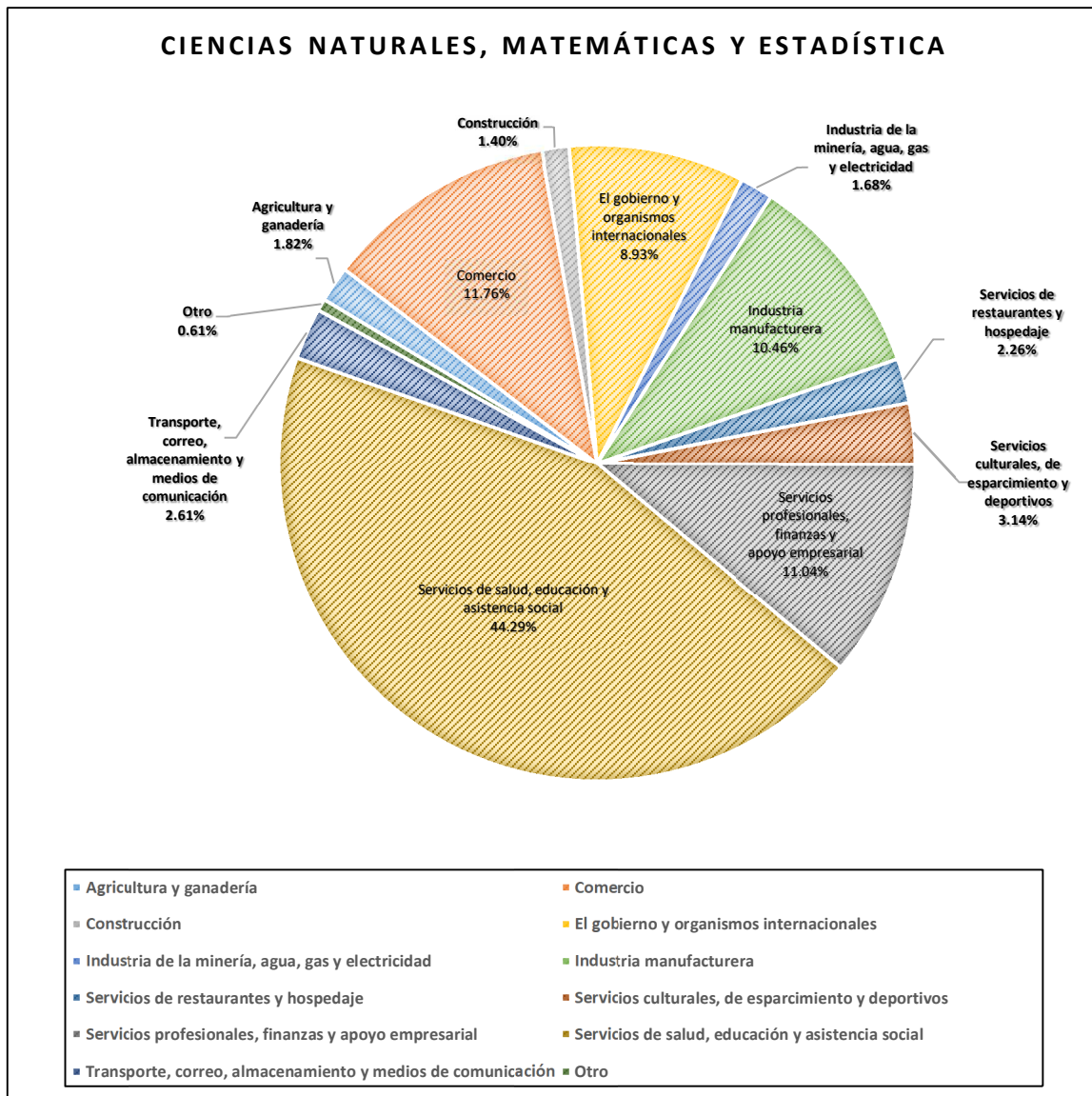


Figura 2. Profesionales de Ciencias naturales, matemáticas y estadística ocupados por sectores de actividad económica

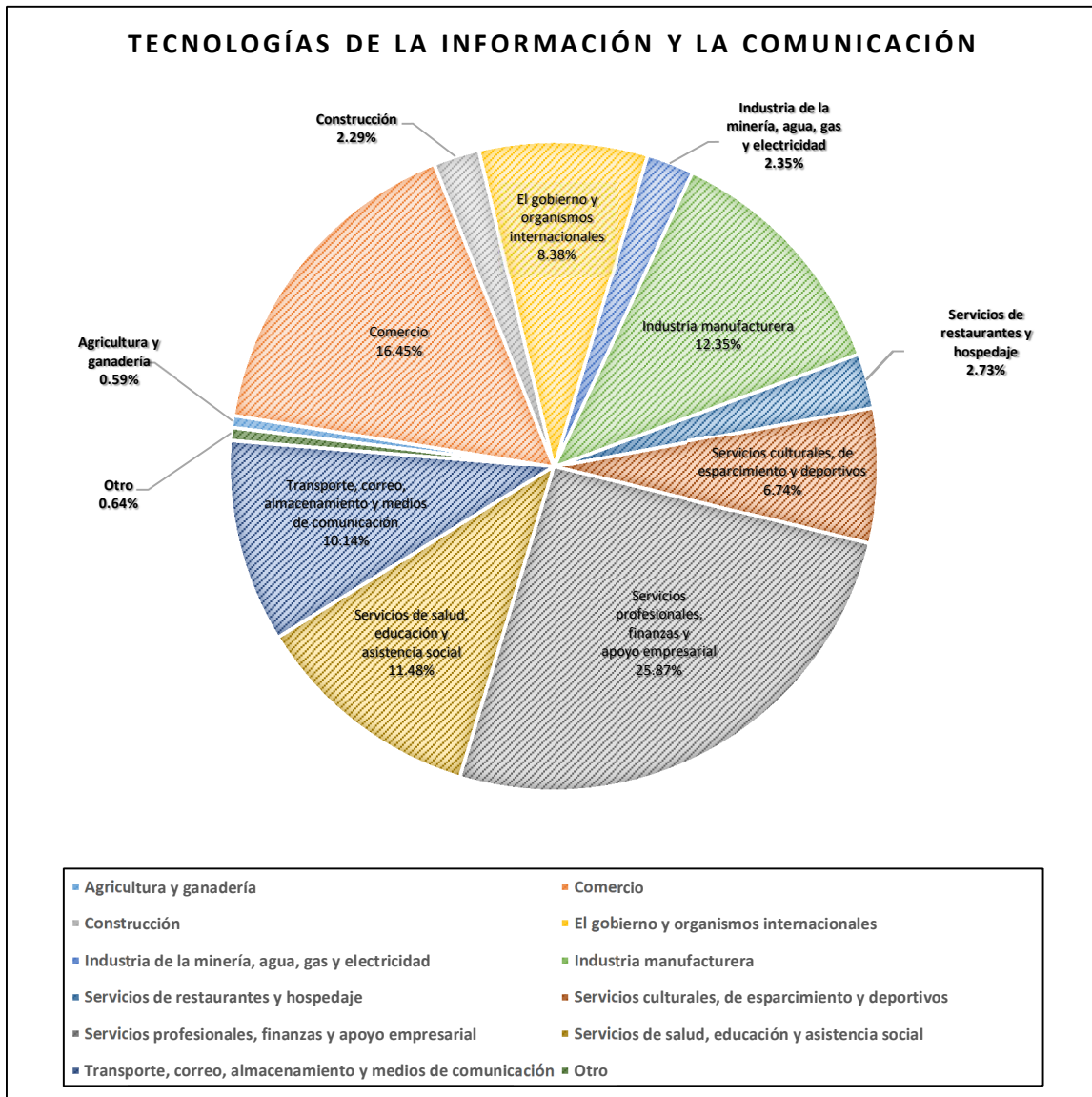


Figura 3. Profesionales de Tecnologías de la información y la comunicación ocupados por sectores de actividad económica

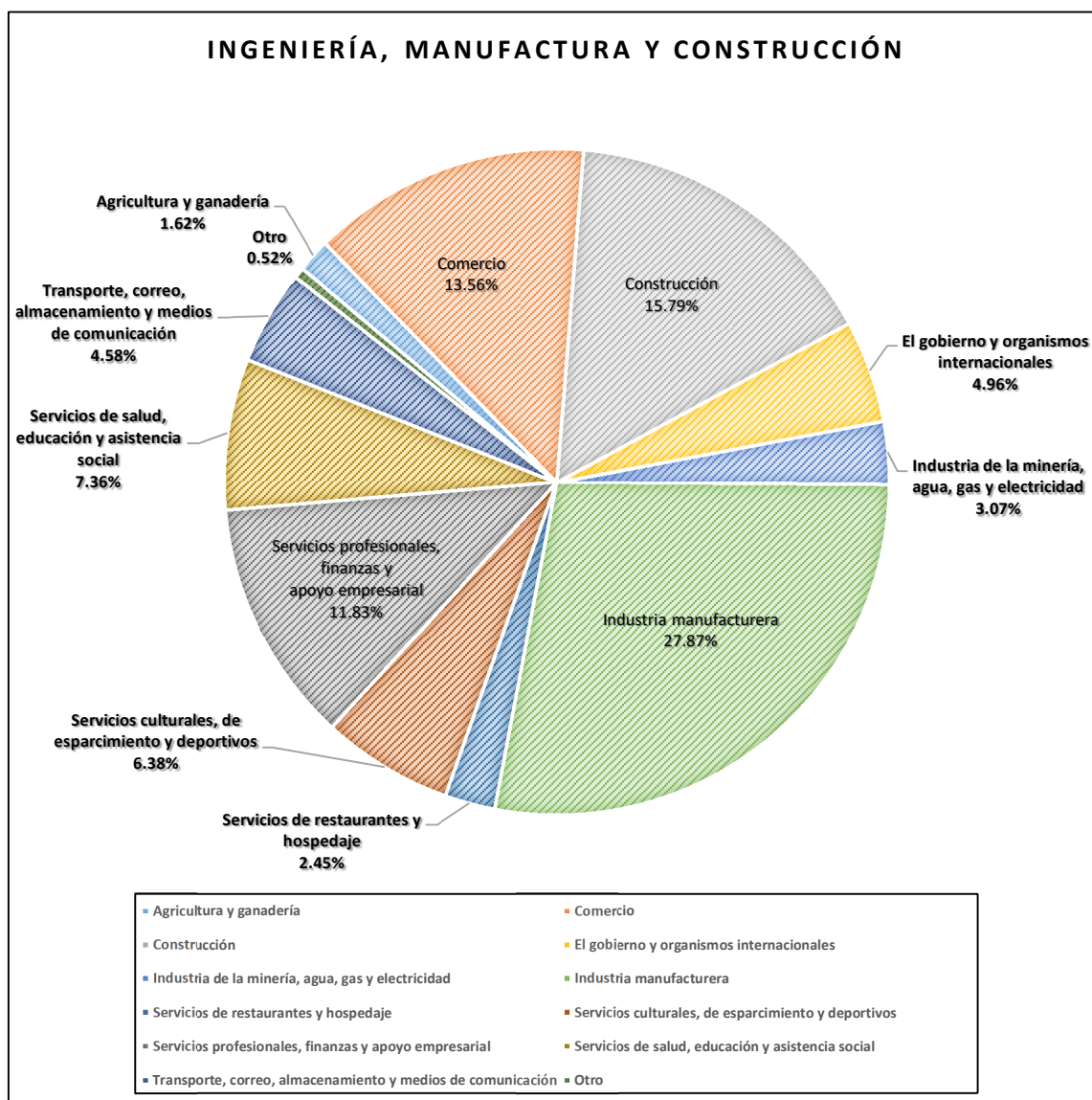


Figura 4. Profesionales de Ingeniería, manufactura y construcción ocupados por sectores de actividad económica

La visualización de datos es útil porque con ella podemos comparar y digerir a simple vista un gran número de datos haciendo más sencillo darse cuenta si existe un patrón o una tendencia. Para el caso de nuestra investigación podemos notar que, para los profesionales de Ciencias naturales, matemáticas y estadística existe un elevado porcentaje de ellos trabajando en el sector de actividad económica Servicios de salud, educación y asistencia social, con un 44.29% de los profesionales ocupados en este sector podemos afirmar que esta es la principal ocupación a la que los estudiantes de las carreras de Matemáticas, Física o Actuaría se dedican.

Sin embargo, contrastando con las otras 2 áreas de profesionales STEM, podríamos preguntarnos si los estudiantes de nuestra facultad no están sobre representados en un sector de actividad económica puesto que para los profesionales de Tecnologías de la información y la comunicación e Ingeniería, manufactura y construcción pareciera que sus proporciones están *mejor* distribuidas, es decir, que dichas proporciones son más homogéneas y no se acumulan con una grande discrepancia en un solo sector de actividad económica. Para responder esta pregunta es plausible realizar una prueba de Chi-cuadrado para saber si las diferencias entre las proporciones de cada sector en los diferentes conjuntos son estadísticamente significativas.

6.4 Prueba χ^2

A continuación, se definirán las hipótesis:

- Hipótesis nula (H_0): No hay diferencia significativa entre los porcentajes de los sectores en los distintos conjuntos, es decir, cualquier diferencia se debe al azar.
- Hipótesis alternativa (H_1): Existe una diferencia significativa, lo que indica que algunos sectores están sobre representados en los conjuntos.

Recapitulando los ocupados por sector que obtuvimos de los datos:

Nombre del Área	Agricultura y ganadería	Comercio	Construcción	El gobierno y organismos Internacionales	Industria de la minería, agua, gas y electricidad	Industria manufacturera	Servicios de restaurantes y hospedaje	Servicios culturales, de esparcimiento y deportivos	Servicios profesionales, finanzas y apoyo empresarial	Servicios de salud, educación y asistencia social	Transporte, correo, almacenamiento y medios de comunicación	Otro
Ciencias naturales, matemáticas y estadística	5257	34069	4067	25862	4872	30300	6546	9093	31974	128264	7555	1764
Tecnologías de la información y la comunicación	4872	135750	18908	69171	19410	101947	22491	55624	213480	94742	83666	5277
Ingeniería, manufactura y construcción	29096	243601	283711	89173	55231	500768	43967	114586	212575	132240	82283	9384

Tabla 4. Número de ocupados de distintas áreas por sector de actividad económica

Bajo la hipótesis nula, esperaríamos que los porcentajes de cada sector sean iguales en todos los conjuntos. Verificamos que se cumplan las condiciones de aplicabilidad:

- $O_{ij} \in \mathbb{N}$: Las entradas en la tabla son frecuencias absolutas de personas ocupadas.
- $\mathbb{P}(X_i = x, X_j = y) = \mathbb{P}(X_i = x)\mathbb{P}(X_j = y), i \neq j$: Se asume muestreo aleatorio simple como registro agregado sin repeticiones, estándar en datos provenientes de censos o encuestas como las del INEGI.
- $n = \sum \sum O_{ij} \gg 30$: Aunque no se especifica una n exacta, se acepta comúnmente que $n \geq 30$ como mínimo empírico para que el estadístico χ^2 se aproxime a su distribución límite.
- $E_{ij} \geq 5$: Revisando la tabla de valores esperados nos encontramos con que no hay celdas con valores esperados pequeños garantizando la validez de la aproximación a la distribución.

Calculamos entonces las frecuencias esperadas bajo la suposición de la hipótesis nula de que no hay relación entre las variables, basándose en el total de observaciones de cada conjunto y la proporción global de cada sector de la siguiente forma:

$$E_{ij} = \frac{(Total\ de\ la\ fila\ i)(Total\ de\ la\ columna\ j)}{Total}$$

Por lo que obtenemos los siguientes valores esperados:

Nombre del Área	Agricultura y ganadería	Comercio	Construcción	El gobierno y organismos Internacionales	Industria de la minería, agua, gas y electricidad	Industria manufacturera	Servicios de restaurantes y hospedaje	Servicios culturales, de esparcimiento y deportivos	Servicios profesionales, finanzas y apoyo empresarial	Servicios de salud, educación y asistencia social	Transporte, correo, almacenamiento y medios de comunicación	Otro
Ciencias naturales, matemáticas y estadística	3901.83	41124.10	30506.96	18323.51	7909.39	62967.86	7261.92	17835.80	45561.49	35337.36	17258.95	1633.84
Tecnologías de la información y la comunicación	11119.02	117191.25	86935.60	52216.47	22539.37	179439.36	20694.28	50826.62	129836.47	100700.80	49182.79	4655.96
Ingeniería, manufactura y construcción	24204.15	255104.65	189243.44	113666.02	49064.23	390607.78	45047.80	110640.58	282631.05	219207.84	107062.25	10135.20

Tabla 5. Valores esperados para la prueba de Chi-cuadrado de los ocupados por sector de actividad económica

Finalmente:

$$\alpha = 0.05$$

$$v = (3 - 1)(12 - 1) = 22 \text{ grados de libertad}$$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx 625858.75 > 33.92 \approx F^{-1}(1 - \alpha, v) = \chi_{crítico}^2$$

Rechazamos así la hipótesis nula. Esto significa que hay evidencia estadística significativa para afirmar que la distribución de los porcentajes entre los diferentes sectores en los tres conjuntos no es igual y que precisamente algunas categorías, como Servicios de salud, educación y asistencia social en el área de Ciencias naturales, matemáticas y estadística, están sobre representadas, es decir, que la distribución de los ocupados en los distintos sectores de actividad económica no es uniforme o aleatoria entre las 3 áreas de profesionales STEM.

6.5 Ingresos por percentiles y promedio

Los ingresos pueden ser un indicador importante en la evaluación de una decisión de carrera, no son el único factor que se debe considerar al determinar si se ha hecho una buena elección de área de estudio o de trabajo pero no puede negarse la importancia y el atractivo que este factor ofrece al desarrollo de una carrera profesional, muchas veces siendo este el factor por el cual un estudiante o un graduado cambia o toma otro camino en su carrera profesional al que originalmente había seleccionado.

Al considerar los percentiles de ingreso, el percentil 25 representa el ingreso por debajo del cual se encuentra el 25% de la población, el percentil 50 (o la mediana) es el ingreso que divide a la población en dos mitades iguales, y el percentil 75 es el ingreso por debajo del cual se encuentra el 75% de la población. El percentil 100, por lo tanto, marca el ingreso más alto, superando al resto de la población.

En una tabla de ingresos, el ingreso en el percentil 100 no solo refleja el tope de los ingresos, sino que también puede ser útil para identificar desigualdades económicas, especialmente cuando se compara con los percentiles inferiores como el 25, 50 y 75.

El ingreso percentil 100 en una tabla de distribución de ingresos representa el ingreso máximo dentro del grupo analizado, lo cual significa en otras palabras que no hay nadie en el grupo que gane más que este valor. Es el punto más alto de la distribución de ingresos, indicando el límite superior del ingreso en la población o el segmento analizado.

En la gráfica proporcionada, se muestran los ingresos correspondientes a los percentiles 25, 50, 75 y 100 para los tres grupos diferentes. Los percentiles nos indican cómo se distribuyen los ingresos dentro de cada área.

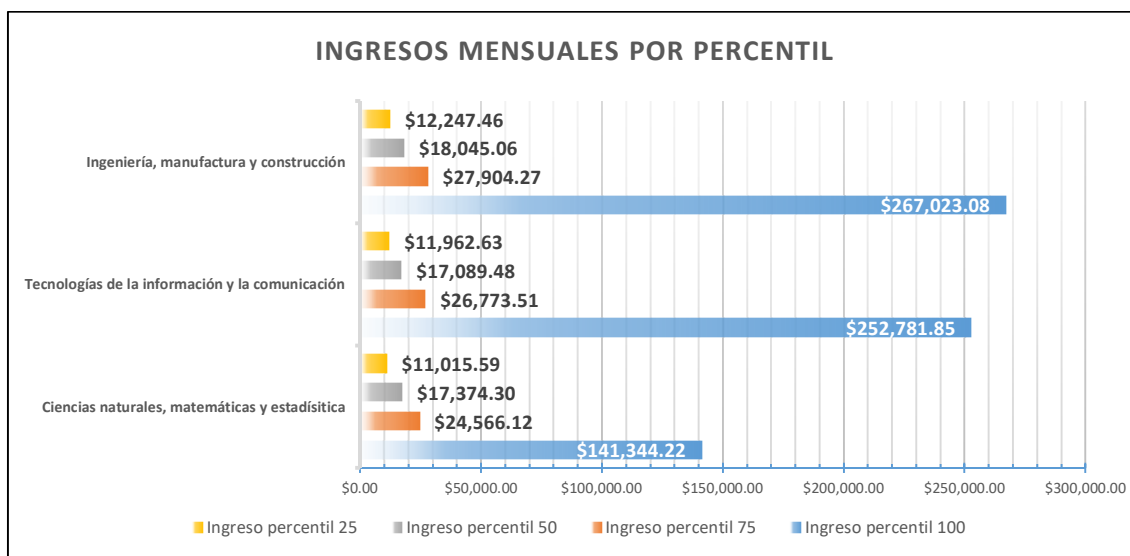


Figura 5. Ingresos correspondientes a los percentiles 25, 50, 75 y 100 para tres áreas diferentes: Ciencias naturales, matemáticas y estadística; Tecnologías de la información y Comunicación e Ingeniería, manufactura y construcción.

La distancia entre los percentiles 75 y 100 puede ser vista como una medida de la dispersión de ingresos. Una mayor distancia indica una mayor dispersión, que significa que dentro del grupo de los mejor pagados, hay una élite aún más reducida con ingresos muchos mayores. En términos de desigualdad para cada una de las áreas:

- *Ciencias naturales, matemáticas y estadística:* Aunque se presenta una diferencia notable entre el ingreso en el percentil 100 (\$141,344.22) y el ingreso en el percentil 75 (\$24,566.12), lo cual implica cierta desigualdad interna entre los mejores pagados, ésta es menor en comparación con las otras áreas (dispersión de \$116,778.10).
- *Tecnologías de la información y la comunicación:* La desigualdad también es notable aquí, con el ingreso en el percentil 100 (\$252,781.85) distanciándose mucho del ingreso en el percentil 75 (\$26,773.51). Al igual que en las ciencias naturales, existe un grupo pequeño con ingresos muy elevados pero es notable que la brecha es mucho mayor (dispersión de \$226,008.34).
- *Ingeniería, manufactura y construcción:* Esta área muestra la mayor desigualdad entre las tres. El ingreso en el percentil 100 (\$267,023.08) es significativamente mayor que el ingreso en el percentil 75 (\$27,904.27), lo que indica que los ingresos más altos están mucho más alejados de la mayoría económica (dispersión de \$239,118.81).

Al observar el crecimiento de los ingresos de un percentil a otro, podemos ver que el salto de ingresos entre los percentiles 75 y 100 es muy pronunciado en todas las áreas, lo que sugiere que los ingresos están altamente sesgados hacia el extremo superior.

La mediana de ingresos (percentil 50) proporciona el ingreso típico y divide la población en dos mitades: una mitad gana más y la otra mitad gana menos que ese valor. Los ingresos medianos son más altos en Ingeniería, manufactura y construcción (\$18,045.06) en comparación con las otras dos áreas, lo que sugiere que, en general, los trabajadores en este campo tienden a tener ingresos más altos.

Estas observaciones pueden ser indicativas de la existencia de una minoría de individuos con ingresos extremadamente altos dentro de

cada área, lo que contribuye a un mayor nivel de desigualdad de ingresos. Este fenómeno es común en muchas industrias y puede ser un reflejo de factores como la concentración de habilidades altamente especializadas, diferencias en la negociación salarial, o la acumulación de roles de alta dirección en el extremo superior de la distribución de ingresos.

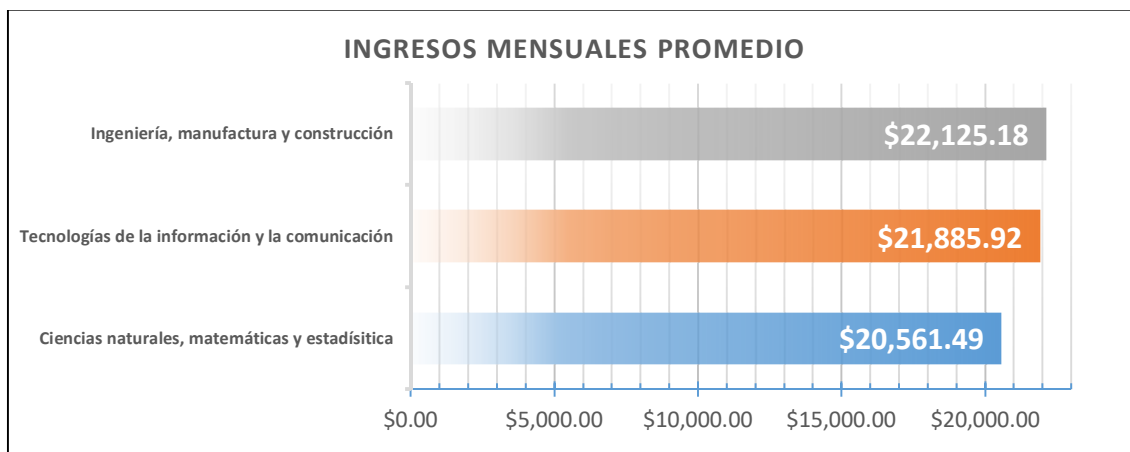


Figura 6. Ingresos promedio para tres áreas diferentes: Ciencias naturales, matemáticas y estadística; Tecnologías de la información y Comunicación e Ingeniería, manufactura y construcción.

Ingeniería, manufactura y construcción tiene el ingreso promedio más alto (\$22,125.18), seguido por Tecnologías de la información y la comunicación (\$21,885.92), y luego Ciencias naturales, matemáticas y estadística (\$20,561.49). Un ingreso promedio más alto podría indicar que, en general, las personas en ese grupo tienden a ganar más, pero este número puede estar inflado por ingresos extremadamente altos en el extremo superior de la escala, es por ello que la mediana es menos susceptible a valores atípicos y proporciona una mejor idea del ingreso típico.

Ingeniería, manufactura y construcción tiene el ingreso más alto en el percentil 100 (\$267,023.08), seguido muy de cerca por Tecnologías de la información y la comunicación (\$252,781.85). Ciencias naturales, matemáticas y estadística tiene un ingreso en el percentil 100 significativamente más bajo (\$141,344.22). Esto nos dice qué tan alto puede esperar llegar en términos de ingresos cada grupo con base en los datos.

Si bien el área de Ciencias naturales, matemáticas y estadística presenta una menor desigualdad en la distribución de ingresos, también es importante señalar un aspecto preocupante: los ingresos del percentil 100 en las áreas de Tecnologías de la información y comunicación e Ingeniería, manufactura y construcción superan con creces a los del mismo percentil en Ciencias naturales, matemáticas y estadística, llegando a casi duplicarlos. Es preocupante porque comparten fundamentos científicos y técnicos similares sugiriendo que, aunque la desigualdad interna sea menor en el grupo de Ciencias naturales, matemáticas y estadística, también significa una menor retribución económica en los niveles más altos, que podría plantearnos interrogantes sobre el reconocimiento y valorización del trabajo en dicho grupo en comparación con otros con bases comparables.

Tomando en cuenta la compensación económica y considerando tanto la mediana, el promedio, así como los ingresos máximos, Ingeniería, manufactura y construcción parece ser la mejor área, ya que lidera en esas categorías. Esto sugiere que, no solo hay una posibilidad de ganar ingresos extremadamente altos, sino que incluso el ingreso típico es el más alto entre las tres áreas.

Tecnologías de la información y la comunicación se encuentra en una posición intermedia, con un buen ingreso mediano apenas por debajo de Ciencias naturales, matemáticas y estadística pero con la posibilidad de ingresos muy altos.

Ciencias naturales, matemáticas y estadística tiene el ingreso promedio y los percentiles más bajos en general y, aunque el ingreso máximo es sustancial, la diferencia entre el ingreso máximo y los otros percentiles es menos pronunciada, lo que sugiere una desigualdad menor entre los otros grupos, pero hacia ingresos más bajos.

6.6 Retorno Sobre la Inversión y Acceso a Empleo de Calidad

El Retorno Sobre la Inversión (RSI) es un cálculo financiero que se utiliza para determinar la rentabilidad de una inversión, dentro de las finanzas a mayor riesgo, debería esperarse mayor retorno [23]. Para nuestro contexto, el RSI sería un indicador del beneficio económico que se puede obtener en relación con el costo de invertir en esa área. Por ejemplo, un RSI más alto indicaría que por cada unidad monetaria invertida en la educación de un campo específico, hay un retorno mayor en términos de ingresos futuros esperados lo cual la convierte en una medida de la eficiencia de la inversión.

Nombre del Área	Costo promedio privada	Retorno sobre inversión privada	Costo promedio pública	Retorno sobre inversión pública
Ciencias naturales, matemáticas y estadística	\$668,563	4.90%	\$32,049	12.90%
Tecnologías de la información y la comunicación	\$351,279	6.90%	\$34,866	12.90%
Ingeniería, manufactura y construcción	\$381,182	6.70%	\$31,748	13.20%

Tabla 6. Costos y Retorno Sobre Inversión de distintas áreas.

La Probabilidad de Acceder a un Empleo de Calidad (PAEC) se define como la probabilidad de que los individuos encuentren un empleo que se considera de "calidad". Esto normalmente incluiría empleo formal, con un ingreso adecuado y condiciones de trabajo estables.

Nombre del Área	Total	Desocupados	Disponibles	No disponible	Ocupado	Formal	Informal	Desanimado
Ciencias naturales, matemáticas y estadística	421,047	14,480	15,507	101,437	289,623	216,200	73,423	6,035
Tecnologías de la información y la comunicación	975,622	27,554	27,336	95,394	825,338	668,482	156,856	14,138
Ingeniería, manufactura y construcción	2,264,809	92,984	62,388	312,822	1,796,615	1,416,985	379,630	46,458

Tabla 7. Totales, desocupados, disponibles, no disponibles, ocupados, formales, informales, y desanimados de distintas áreas.

La PAEC se calcula como uno menos la probabilidad de que un egresado de cierta carrera se encuentre desocupado, trabajando en el sector informal o formando parte de la población desanimada:

$$PAEC = 1 - riesgo = 1 - \frac{Desocupados + Informales + Desanimados}{Total}$$

Nombre del Área	Probabilidad de Acceder a un Empleo de Calidad	Riesgo	Tasa de desocupados	Tasa de informales	Tasa de desanimados
Ciencias naturales, matemáticas y estadística	77.69%	22.31%	4.76%	25.35%	5.16%
Tecnologías de la información y la comunicación	79.65%	20.35%	3.23%	19.01%	11.52%
Ingeniería, manufactura y construcción	77.08%	22.92%	4.92%	21.13%	12.38%

Tabla 8. Probabilidad de Acceder a un Empleo de Calidad, riesgo, tasa de desocupados, tasa de informales y tasa de desanimados para distintas áreas.

Con los datos de inversión pública y privada es posible realizar un gráfico del RSI de cada área versus el riesgo que le precede estudiar el área correspondiente y trazar una línea de Riesgo/Rendimiento promedio (RRP) que une al origen con el punto conformado por el RSI y riesgo promedio de todas las áreas con el propósito de establecer una referencia objetiva del equilibrio entre beneficio y riesgo.

Se calculó la distancia perpendicular de cada punto (área) con respecto a la línea RRP. Estas distancias permiten clasificar las áreas según la eficiencia relativa de la inversión educativa. A partir de estas distancias, se determinaron dos umbrales: una línea verde, que marca la media de las distancias positivas, y una línea roja, que indica la media de las distancias negativas absolutas.

Este procedimiento fue realizado de manera independiente tanto para el RSI público como para el privado, de forma que fuera posible comparar cómo varía la rentabilidad relativa de las áreas en función del tipo de inversión educativa.

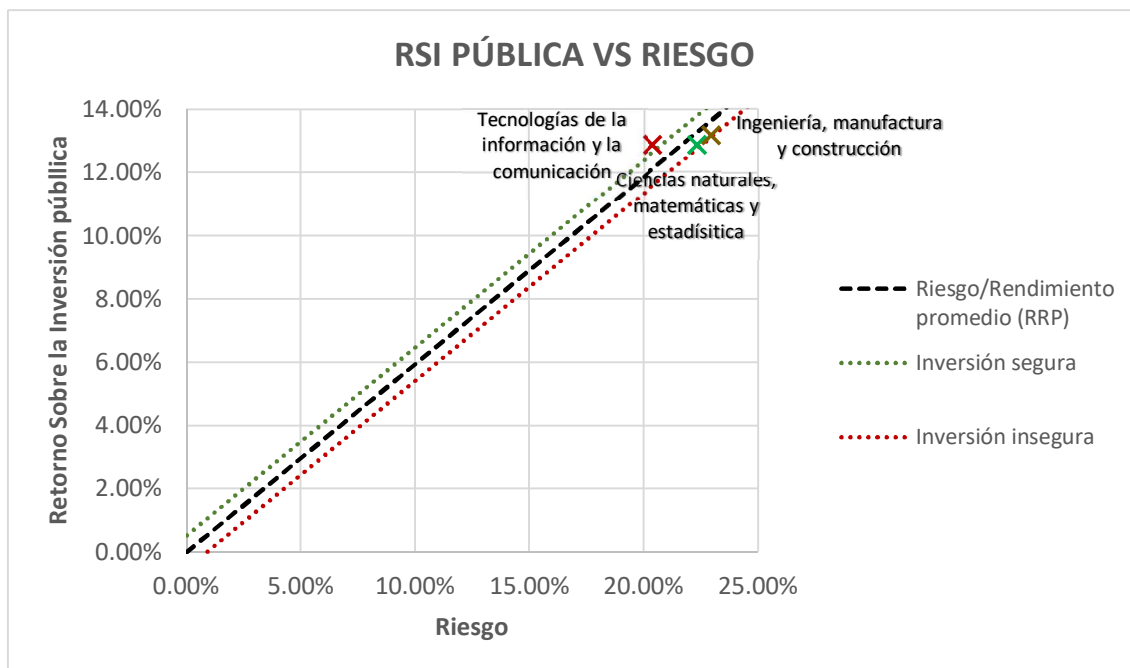


Figura 7. Retorno Sobre la Inversión vs Riesgo por estudiar carreras de cada área en universidades públicas.

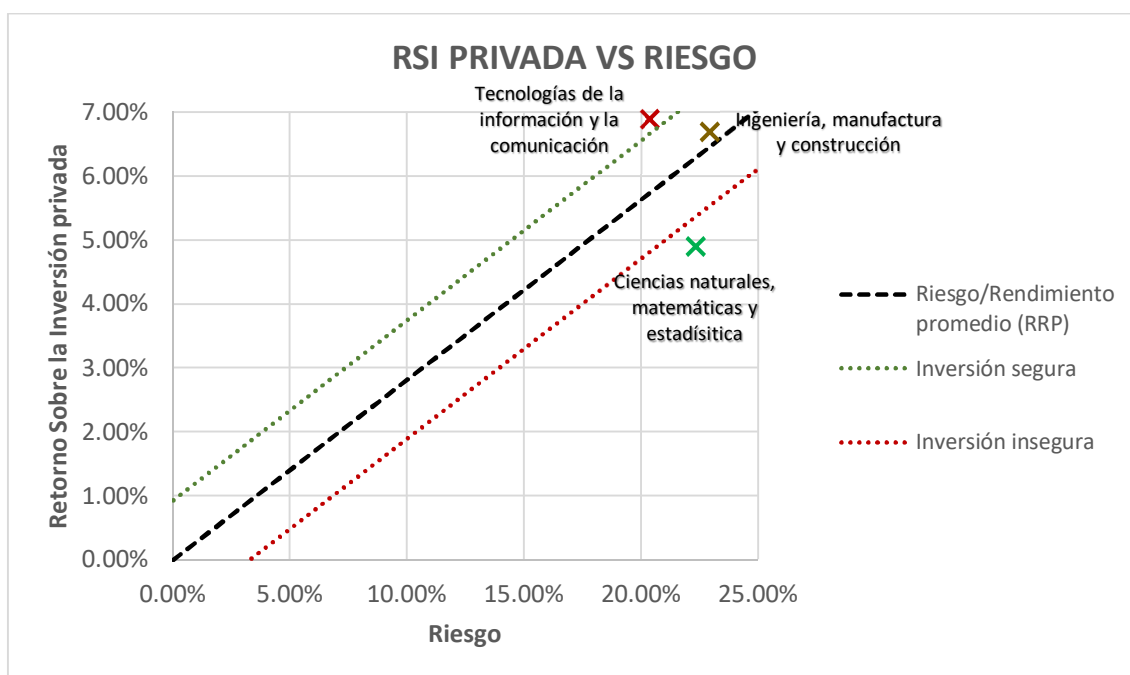


Figura 8. Retorno Sobre la Inversión vs Riesgo por estudiar carreras de cada área en universidades privadas.

Las áreas ubicadas por encima de la línea verde son aquellas con un desempeño destacado, ya que presentan un RSI alto en relación con un riesgo bajo; las que se encuentran entre la línea RRP y la línea verde son positivas, aunque con un retorno menos pronunciado. Por otro lado, las áreas por debajo de la línea RRP pero por encima de la línea roja se

consideran inversiones de riesgo moderado, mientras que aquellas más alejadas negativamente (por debajo de la línea roja) representan una inversión educativa con baja eficiencia relativa, combinando un alto riesgo con un retorno limitado.

En la gráfica de RSI pública, las tres áreas analizadas se ubican muy cercanas entre sí y dentro de una zona de rendimiento positivo aceptable, aunque con niveles de riesgo similares. El área en Tecnologías de la información y la comunicación se posiciona ligeramente por encima de la línea verde, lo que la clasifica como una inversión segura y eficiente, mientras que Ingeniería, manufactura y construcción y Ciencias naturales, matemáticas y estadística se mantienen cerca de la media apenas siendo inversiones seguras, sin evidenciar desventajas claras entre ellas.

Por el contrario, en la gráfica de RSI privada, se observa una mayor diferencia, aunque Tecnologías de la información y la comunicación mantiene una posición favorable, Ciencias naturales, matemáticas y estadística se ubica por debajo de la línea roja, lo que sugiere una relación poco eficiente entre el costo privado y retorno futuro, clasificándola como una inversión insegura en comparación con las otras dos áreas. Ingeniería, aunque con mayor riesgo, se mantiene cerca del promedio.

Estas gráficas evidencian que, si bien las tres áreas STEM presentan riesgos laborales relativamente similares, el tipo de institución educativa (pública o privada) influye significativamente en la eficiencia de la inversión. Específicamente, la inversión pública mejora el rendimiento relativo de las carreras menos rentables en el sector privado, lo cual tiene implicaciones importantes para la planeación educativa y la toma de decisiones de sus estudiantes.

Conclusiones

En el presente proyecto de investigación se ha identificado y argumentado la necesidad de integrar el análisis de datos como área terminal en el currículo de las carreras de Física en la Facultad de Ciencias Físico-Matemáticas. Este esfuerzo responde a una exigencia latente en el mercado laboral y a la creciente importancia de las competencias en análisis de datos en el contexto de la industria y la academia.

Vivimos en un mundo regido por datos, donde las decisiones estratégicas en sectores como la salud, tecnología, finanzas y educación dependen de la correcta interpretación y manejo de grandes volúmenes de información. Los egresados de física, si bien cuentan con una importante formación matemática y analítica, no disponen de herramientas técnicas específicas para saber cómo manejar los datos colocándolos en una posición desventajosa frente a otros profesionales de áreas STEM que han integrado el análisis de datos en su formación académica.

Las tendencias globales y nacionales evidenciadas en este estudio muestran que el análisis de datos no es solo una habilidad deseable, sino un requisito esencial para competir en el mercado laboral actual. Los sectores de Tecnologías de la Información y la Comunicación e Ingeniería, Manufactura y Construcción cuentan con mayores ingresos promedio y oportunidades de empleo más diversificadas, contrastando con el área de Ciencias Naturales, Matemáticas y Estadística, que presenta limitaciones significativas en su alcance laboral y salarial.

La encuesta aplicada a estudiantes de la facultad revela un interés generalizado en el aprendizaje de análisis de datos, con más del 85% de los encuestados considerando necesario aumentar la oferta de materias en esta área. Además, un 78% mostró disposición para cursar materias adicionales relacionadas con análisis de datos aplicados a su campo de estudio asegurando un alto grado de aceptación por parte de los estudiantes para esta propuesta.

Los egresados de Ciencias Naturales, Matemáticas y Estadística están tienen una mayor presencia relativa en sectores como salud, educación y asistencia social, lo cual contrasta con la distribución más diversificada que presentan sus pares egresados de Tecnologías de la Información y la Comunicación e Ingeniería, Manufactura y Construcción, quienes se encuentran laborando en otros sectores de alto valor económico.

La integración del análisis de datos en el currículo no solo beneficiará la empleabilidad de los egresados, sino que también abrirá nuevas oportunidades interdisciplinarias y fomentará la innovación en sectores emergentes. Los físicos, con su formación rigurosa, tienen el potencial de destacar en campos como la ciencia de datos, machine learning, modelado predictivo y análisis estadístico, áreas que demandan precisamente el tipo de competencias que esta investigación busca desarrollar alineando también a la Facultad de Ciencias Físico-Matemáticas de la BUAP con las tendencias educativas globales, para contribuir al prestigio institucional y fortaleciendo su papel como referente en la formación de profesionales STEM.

Tabla de contenidos

Tabla 1. Tabla operacional de valores y variables.	47
Tabla 2. Resultados de la aplicación de encuesta.....	54
Tabla 3. Métricas del repositorio del IMCO para el análisis.....	57
Tabla 4. Número de ocupados de distintas áreas por sector de actividad económica	61
Tabla 5. Valores esperados para la prueba de Chi-cuadrado de los ocupados por sector de actividad económica	62
Tabla 6. Costos y Retorno Sobre Inversión de distintas áreas.....	68
Tabla 7. Totales, desocupados, disponibles, no disponibles, ocupados, formales, informales, y desanimados de distintas áreas.....	68
Tabla 8. Probabilidad de Acceder a un Empleo de Calidad, riesgo, tasa de desocupados, tasa de informales y tasa de desanimados para distintas áreas.	69
Figura 1. Diagrama de Venn que muestra las relaciones entre las distintas áreas que dan lugar a la analítica de datos, donde ocurre el análisis de datos	30
Figura 2. Profesionales de Ciencias naturales, matemáticas y estadística ocupados por sectores de actividad económica	58
Figura 3. Profesionales de Tecnologías de la información y la comunicación ocupados por sectores de actividad económica.....	59
Figura 4. Profesionales de Ingeniería, manufactura y construcción ocupados por sectores de actividad económica	60
Figura 5. Ingresos correspondientes a los percentiles 25, 50, 75 y 100 para tres áreas diferentes: Ciencias naturales, matemáticas y estadística; Tecnologías de la información y Comunicación e Ingeniería, manufactura y construcción.....	64
Figura 6. Ingresos promedio para tres áreas diferentes: Ciencias naturales, matemáticas y estadística; Tecnologías de la información y Comunicación e Ingeniería, manufactura y construcción.	66
Figura 7. Retorno Sobre la Inversión vs Riesgo por estudiar carreras de cada área en universidades públicas.....	70
Figura 8. Retorno Sobre la Inversión vs Riesgo por estudiar carreras de cada área en universidades privadas.....	70

Referencias

ADIP CDMX. (s.f.). Gobierno Digital. Recuperado de <https://www.adip.cdmx.gob.mx/centros/gobierno-digital>

[23]

Bodie, Z., Kane, A., & Marcus, A. J. (2021). Investments (11th ed.). McGraw-Hill Education.

[6]

Bureau of Labor Statistics, Departamento de Trabajo de los Estados Unidos. (s.f.). Matemáticos y estadísticos. Manual de Perspectivas Ocupacionales. Recuperado de <https://www.bls.gov/ooh/math/mathematicians-and-statisticians.htm>

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4), 1165-1188.

Data Science for Social Good Fellowship. (2022). 2022 Fellows & Mentors. Recuperado de <https://www.dssgfellowship.org/people/2022-fellows-mentors/>

[13]

Davenport, T. H., & Dyché, J. (2013). Big Data in Big Companies. International Institute for Analytics.

[1]

Departamento de Comercio de los Estados Unidos. (2017). Empleos STEM: Actualización 2017. Recuperado de <https://www.commerce.gov/data-and-reports/reports/2017/03/stem-jobs-2017-update>

Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64-73.

Dun & Bradstreet. (s.f.). Perfil de la empresa Deep Dive Data Science SAPI de CV. Recuperado de https://www.dnb.com/business-directory/company-profiles/deep_dive_data_science_sapi_de_cv.eb3c05c38028d02f5d67892983578114.html

El Financiero. (2022, enero 28). Clark, 'el que hacía datos' antes de que su análisis fuera popular. Recuperado de <https://www.elfinanciero.com.mx/opinion/maria-scherer-ibarra/2022/01/28/clark-el-que-hacia-datos-antes-de-que-su-analisis-fuera-popular/>

El Universal. (2016, 20 de diciembre). Detective de datos. Tras el oro escondido en los números. Recuperado de <https://www.eluniversal.com.mx/articulo/nacion/2016/12/20/detective-de-datos-tras-el-oro-escondido-en-los-numeros>

Elmasri, R., & Navathe, S. B. (2015). *Fundamentals of Database Systems*. Pearson.

[7]

Etzkowitz, Henry (2008). *The triple helix: university-industry-government innovation in action*. Routledge.

[21]

Facultad de Ciencias Físico Matemáticas, BUAP. (2024). Cuarto Informe de Labores 2024. Recuperado de https://www.fcfm.buap.mx/assets/docs/nosotros/autoridades/direccion/informes/4to_Informe_2024.pdf

[14]

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.

Gandomi, A., & Haider, M. (2015). Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, 35(2), 137–144.

García, C. M., & Cook, S. (2013). *STEM (Science, Technology, Engineering, and Mathematics) Workforce Trends & Policy Considerations*. Nova Science Publishers, Inc.

[10]

Guan, J. C., Yam, R. C. M., & Mok, C. K. (2005). Collaboration between industry and research institutes/universities on industrial innovation in Beijing,

China. *Technology Analysis and Strategic Management*, 17(3), 339–353.

<https://doi.org/10.1080/09537320500211466>

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

[16]

Hayashi, C. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. In: Hayashi, C., Yajima, K., Bock, HH., Ohsumi, N., Tanaka, Y., Baba, Y. (eds) *Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Tokyo. https://doi.org/10.1007/978-4-431-65950-1_3

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (6ª ed.). México: McGraw Hill.

[5]

Hernández-Cordero, L. L., & Alva-Sánchez, H. (2022). Estudio sobre la situación laboral actual de egresados de las carreras de física e ingeniería física en México. *Revista Mexicana de Física*, 68(3), 1–9. Recuperado de <https://rmf.smf.mx/ojs/index.php/rmf/article/view/6007/6332>

[12]

Inform. (2013). Data Science and Prediction. Communications of the ACM, 56(12), 64-73. Recuperado de <https://pubsonline.informs.org/doi/10.1287/LYTX.2013.02.07/full/>

[11]

Inmon, W. H. (2005). Building the Data Warehouse. Wiley.

Instituto Mexicano para la Competitividad (IMCO). (2023a). Compara carreras 2023. Recuperado de https://imco.org.mx/wp-content/uploads/2023/08/VF_Compara_Carreras-1.xlsx

Instituto Mexicano para la Competitividad (IMCO). (2023b). Metodología - Compara Carreras. Recuperado de <https://imco.org.mx/comparacarreras/metodologia>

Instituto Mexicano para la Competitividad A.C. Compara Carreras. Recuperado de https://comparacarreras.imco.org.mx/?gad_source=1&gclid=EAlaIqo bChMImva7x4rWiQMvDrUAR0blwp-EAAYASABEgLhDPD_BwE#Inicio

[3]

Instituto Nacional de Estadística y Geografía (INEGI). (2020). Censo 2020 Resultados complementarios [PDF]. Recuperado de https://www.inegi.org.mx/contenidos/programas/ccpv/2020/doc/Censo2020_Resultados_complementarios_EUM.pdf

[22]

Instituto Nacional de Estadística y Geografía. (2020). Encuesta Nacional de Ocupación y Empleo (ENOE): Nueva estructura de clasificación de ocupaciones. Recuperado de https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825086664.pdf

INEGI. (s.f.). Normateca Institucional. Recuperado de <https://sc.inegi.org.mx/Normateca2010/menuNormateca.jsp?codigo=0301&opc=1>

Intriago Alcívar, G. C., Camacho Tovar, G. L., Sánchez Soto, M. A., Carpio Vera, D. A., & Mendiburu Rojas, A. F. (2019). Metodología de la investigación educativa: retos y perspectivas (pp. 31-51). Las Tunas: Edacun. Recuperado de <https://elibro.net/es/ereader/ieue/151754>

ITAM. (2013, 23 de mayo). Alejandro Maza Ayala, premio TR35 de la revista MIT Technology Review. Recuperado de <https://eventos.itam.mx/es/3/noticias/2013/05/23/alejandro-maza-ayala-premio-tr35-de-la-revista-mit-technology-review>

[19]

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning (2nd ed.). Springer.

[15]

Laursen, G. H. N., & Thorlund, J. (2010). Business Analytics for Managers: Taking Business Intelligence Beyond Reporting. Wiley.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT Sloan Management Review, 52(2), 21.

[9]

Leydesdorff, L. (2009). The Communication of Meaning and the Structuration of Expectations: Giddens' "Structuration Theory" and Luhmann's "Self-Organization." Journal of the American Society for Information Science and Technology, 60(4), 845–857. Recuperado de <https://www.leydesdorff.net/arist09/arist09.pdf>

Leydesdorff, L., & Van Den Besselaar, P. (2006). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. Journal of the American Society for Information Science and Technology, 57(11), 1450-1462. <https://doi.org/10.1080/09537320500211466>

LinkedIn. (s.f.). Perfil de Alejandro Maza. Recuperado de <https://www.linkedin.com/in/mazaalejandro/>

LinkedIn. (s.f.). Perfil de Andrea Fernández Conde. Recuperado de <https://mx.linkedin.com/in/andrea-fernandez-conde>

LinkedIn. (s.f.). Perfil de Manuel Aragonés. Recuperado de <https://www.linkedin.com/in/aragonesm/>

Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. John Murray.

Meneses, J. (Coord.), Rodríguez Gómez, D., & Valero S. (2019). Investigación educativa: una competencia profesional para la intervención (pp. 71-75). Barcelona: Editorial UOC. Recuperado de <https://elibro.net/es/ereader/ieu/124686>

Mundo ITAM. (2021). ¿Cómo ha progresado la Ciencia de Datos en México? Recuperado de <https://mundoitam.com/2021/11/03/como-ha-progresado-la-ciencia-de-datos-en-mexico/>

[2]

Oficina del Censo de los Estados Unidos. (2014). La Oficina del Censo informa que la mayoría de los graduados universitarios en STEM no trabajan en ocupaciones STEM. Recuperado de <https://www.census.gov/newsroom/archives/2014-pr/cb14-130.html>

[18]

Parmenter, D. (2015). Key Performance Indicators: Developing, Implementing, and Using Winning KPIs. Wiley.

[4]

Pew Research Center. (2018). Mujeres y hombres en STEM a menudo en desacuerdo sobre la equidad en el lugar de trabajo. Recuperado de <https://www.pewresearch.org/social-trends/2018/01/09/women->

and-men-in-stem-often-at-odds-over-workplace-equity/ps_2018-01-09_stem_1-09/

PitchBook. (s.f.). Perfil de Alejandro Maza. Recuperado de <https://pitchbook.com/profiles/person/78233-59P#overview>

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), 64-88.

Pozzo, M. I. (2020). Escritura de tesis de posgrado: desde el proyecto hasta la defensa (pp. 31-53). Buenos Aires: Biblos. Recuperado de <https://elibro.net/es/ereader/ieu/153300>

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.

SG.com.mx. (s.f.). Andrea Fernández Conde | SG Buzz. Recuperado de <https://sg.com.mx/buzz/autores/andrea-fernandez-conde>

SG.com.mx. (s.f.). Andrea Fernández Conde. Recuperado de <https://sg.com.mx/dataday/speakers/andrea-fernandez-conde/>

SG.com.mx. (s.f.). Liliana Millán. Recuperado de <https://sg.com.mx/dataday/speakers/liliana-millan/>

Simon, H. A. (1969). *The Sciences of the Artificial*. MIT Press.

[8]

STI/OECD (1999), *University Research in Transition*, OECD, Paris.

The Org. (s.f.). Alejandro Maza - Fundador y Director General at OPI Analytics.
Recuperado de <https://theorg.com/org/opi-analytics/org-chart/alejandro-maza>

The University of Chicago Harris School of Public Policy. (2019, junio 3). Data and Disappearances: Manuel Aragonés Uses Harris Toolkit on Mexico's Missing Persons Crisis. Recuperado de <https://harris.uchicago.edu/news-events/news/data-and-disappearances-manuel-aragones-uses-harris-toolkit-mexicos-missing>

[20]

Turban, E., Sharda, R., Delen, D., & King, D. (2011). Business Intelligence: A Managerial Approach. Pearson.

UNECE. (2023). UNECE Expert Meeting on Dissemination and Communication of Statistics 2023. Recuperado de <https://unece.org/info/events/event/376499>

Universidad Adolfo Ibáñez. (s.f.). Liliana Millán - Escuela de Gobierno. Recuperado de <https://gobierno.uai.cl/profesor-externo-uai/liliana-millan/>

[17]

Watson, H. J. (2006). Business Intelligence – Past, Present, and Future. Communications of the Association for Information Systems, 25, 487–510.

World Economic Forum. (s.f.). Alejandro Maza. Recuperado de <https://www.weforum.org/people/alejandro-maza/>