



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

# RUTAS ESTRATÉGICAS

PROYECTO FINAL

P R E S E N T A

Covarrubias Colindres José de Jesús  
Epitacio Allende Irvin Horacio  
Esquivel Osorio Flor Amanda  
García Rodríguez María Guadalupe  
Martínez Alonso Daniela Berenice  
Medina Roman Carlos Joaquín  
Pérez Valle Karla Fernanda

Profesor: José Gustavo Fuentes Cabrera





---

# ÍNDICE GENERAL

<b>1. Introducción</b>	<b>1</b>
1.1. Planteamiento del problema . . . . .	2
1.2. Objetivo general . . . . .	2
<b>2. MARCO TEÓRICO</b>	<b>5</b>
2.1. Programación Estructurada . . . . .	5
2.2. Programación Orientada a Objetos (POO) . . . . .	6
2.3. Bases de Datos Relacionales . . . . .	7
<b>3. MANEJO DE RDBMS</b>	<b>9</b>
3.1. Diseños Relacionales . . . . .	9
3.2. Diseños para Minería de Datos . . . . .	10
3.3. Procesos de Extracción-Transformación de Carga (ETL) . . . . .	11
<b>4. INTELIGENCIA DE NEGOCIOS</b>	<b>13</b>
4.1. Introducción . . . . .	13
4.2. Indicadores clave de desempeño (KPI's) . . . . .	13
4.2.1. ¿Por qué los KPI definen el éxito? . . . . .	14
4.2.2. ¿Cómo utilizan los KPI las compañías garantizadas? . . . . .	14
4.2.3. KPI's de marketing . . . . .	14
4.3. Diseño de Tableros . . . . .	15
<b>5. INTELIGENCIA ANALITICA</b>	<b>17</b>
5.1. Metodología del Modelado . . . . .	17
5.1.1. Aprendizaje Automático . . . . .	18
5.1.2. Árboles de Desición . . . . .	18
5.1.3. Regresión . . . . .	19
5.1.4. Redes Neuronales . . . . .	19
5.1.5. Máquinas Vector Soporte . . . . .	19
5.1.6. K vecinos más cercanos . . . . .	19
5.1.7. Clasificador ingenuo de Bayes . . . . .	20
5.2. Ensamble de Modelos . . . . .	20

5.2.1. Bosques Aleatorios . . . . .	20
5.2.2. AdaBoost(Adaptative Boosting), Impulso Adaptativo . . . . .	21
5.2.3. Impulso de Árbol Gradiente . . . . .	21
<b>6. Desarrollo de la aplicación de negocio</b>	<b>23</b>
6.1. MISIÓN . . . . .	23
6.2. VISIÓN . . . . .	23
6.3. MODO DE OPERACIÓN . . . . .	23
6.4. SEGURIDAD ESTUDIANTIL . . . . .	27
6.5. IMPLEMENTACIÓN DE MACHINE LEARNING . . . . .	28
<b>7. ANÁLISIS EXPLORATORIO</b>	<b>29</b>
<b>8. CÓDIGO-REGRESIÓN</b>	<b>33</b>
<b>9. DICCIONARIO DE DATOS</b>	<b>37</b>
<b>10. Conclusiones</b>	<b>41</b>
<b>Bibliografía</b>	<b>43</b>

---

# CAPÍTULO 1

---

## INTRODUCCIÓN

Pudiera parecer banal el propósito de buscar soluciones al modo de transporte que tienen que usar los estudiantes para poder llegar a la universidad. De cualquier modo, gran parte de la sociedad de la Facultad de Estudios Superiores Acatlán hace uso del transporte público y por lo tanto también queda expuesta a otro tipo de factores que van de la mano junto con el uso del transporte público. De aquí nace la inquietud de hacer una propuesta para mejorar el transporte público de los estudiantes de la Facultad de Estudios Superiores Acatlán.

Nuestro interés por el traslado de los estudiantes surge a raíz de ver como muchos alumnos de este plantel carecen de un buen transporte público, ya que la mayoría de las veces este es de un costo muy elevando, además de que la calidad es pésima y sumando a esto muy inseguro y en ocasiones es preferible el no asistir a la escuela que hacer uso de este.

Adentrarnos a investigar la información que va relacionada con esta problemática no fue nada sencillo ya que tan solo el hecho de que los estudiantes proporcionen información personal como municipio/delegación, colonia y calle de residencia no es un dato que cualquiera proporciona con facilidad, a pesar de que se les comentaba que estos datos iban a ser usados únicamente para fines estadísticos. Y también el adentrarnos a la propuesta no es nada sencillo.

Como parte de la misma sociedad de alumnado de la Facultad de Estudios Acatlán reconocemos que en esta se vive una gran inseguridad y más aún en los alrededores, por lo que los traslados son muy complicados y más aún en altas horas de la noche o muy temprano, por lo que al no poder ser indiferentes ante esta situación hemos hecho una propuesta que ayudará a la sociedad de esta facultad y que tras observar el éxito de esta se puede expandir a las demás facultades.

Por nuestro interés y virtud de que cada individuo mantiene diferentes necesidades en cuanto al transporte de su residencia a la facultad y que a pesar de que cada uno puede tomar la decisión de usar el transporte que mejor se le acomode, ofrecemos la posibilidad de hacer ese traslado en un lapso menor de tiempo, y únicamente con alumnado de la misma facultad y de la misma zona, este transporte seguiría una ruta estratégica para que a todo el alumnado de esta zona se le deje lo más cerca posible de su residencia.

A partir de ese mismo interés se plantearon las siguientes problemáticas: ¿Se necesitará mayor trans-

porte para la CDMX o el Estado de México?, ¿Qué tipo de transporte será necesario usar?, ¿Es un negocio rentable?, ¿Será útil para los alumnos?, ¿Cuántas rutas serán necesarias?, ¿Hacia qué tipo de público va dirigida esta propuesta? Esto ocasionó que partiéramos de una detección de necesidades del alumnado, por medio de un cuestionario dirigido a la comunidad de la facultad. Este sirvió de base para hacer la base de las perspectivas y así poder hacer la propuesta de rutas estratégicas.

Esa idea va a necesitar de más que nada de un apoyo económico y una buena difusión al comienzo de esta idea, que con el paso del tiempo este podría ser usado por el mismo personal de la facultad.

El lector encontrará en este trabajo de investigación un enlace de elementos que resaltarán todo el análisis de datos que se hizo para poder llevar a cabo, conforme el texto vaya avanzando se irán explicando como es que se tomaron las decisiones, principalmente sobre la decisión de las rutas estratégicas. Para finalmente poder hacer conclusiones sobre los datos obtenidos y la explotación de estos y así poder hacer una buena estrategia de negocio y no solo eso, lo más importante que es dar solución a una problemática que la gran mayoría de los estudiantes de la Facultad de Estudios Superiores Acatlán se enfrentan día a día.

## 1.1. Planteamiento del problema

En este trabajo de investigación lo que se plantea es dar una solución a la problemática con la que se enfrentan los alumnos de la Facultad de Estudios Superiores Acatlán que es el uso del transporte público y que como es bien sabido el uso de este va de la mano junto con otras cuestiones como en la mayoría de los casos es la inseguridad, que es el principal factor que buscamos disminuir, al menos para los estudiantes de esta facultad. Por lo que fue necesario hacer un levantamiento de información cualitativa respecto a la cantidad de personas que hacen uso del transporte público y como es que se haría una red de transporte para hacer rutas estratégicas. Y esta problemática comienza a ser abordada a partir de un cuestionario en la que se hace un cuestionamiento sobre las principales desventajas que tiene el uso del transporte público y no solo eso sino eliminar estas o en caso de que no sea posible disminuirlo es lo posible. Y a partir de esto es que se desarrolla la idea de rutas estratégicas.

## 1.2. Objetivo general

Ante esta problemática con la que la mayoría de los estudiantes se encuentra buscamos disminuir en medida de lo posible eliminarlas o disminuirlas. La de mayor importancia es la inseguridad, esta quedaría en niveles mínimos ya que de este transporte únicamente comunidad de esta facultad podría hacer uso. Después de esto se busca que el transporte sea de buena calidad, es decir que ya sean autobuses o combis estén en un buen estado, no solo mecánico sino también en lo estético y de esto viene de la mano otro factor a disminuir que es la limpieza se busca que este siempre este limpio. Después de esto se buscó atender el hecho del tiempo de traslado, siento que este sería un transporte relativamente directo, podría tomar rutas alternas mucho más eficientes y así tendríamos la ventaja de reducir de manera muy radical el tiempo de trasladados que en algunos casos considero que este es un factor primordial, ya que en muchos casos este llega a ser superior a los 120 minutos (2 horas). Y este tiempo podría ser invertido en otras cosas de mayor provecho. Otro punto muy importante a tratar es el precio de este transporte.





---

---

# CAPÍTULO 2

---

## MARCO TEÓRICO

### 2.1. Programación Estructurada

Los primeros lenguajes de programación de alto nivel dejaban al programador mucha libertad para definir que estructura le daban al código. Se permitía que el programador alterara el flujo de ejecución del programa a su antojo, algunos lenguajes de programación más antiguos se apoyaban en una sola instrucción para modificar la secuencia de ejecución de las instrucciones mediante una *transferencia incondicional* de su control (con la instrucción *goto*, del inglés "go to", que significa ir .<sup>a</sup>).

A finales de los años sesenta, surgió una nueva forma de programar que reduce a la mínima expresión el uso de la instrucción *goto* y la sustituye por otras más comprensibles. Esta forma de programar se basa en un famoso teorema, desarrollado por Edsger Dijkstra, que demuestra que todo programa puede escribirse utilizando únicamente las tres estructuras básicas de control que se mencionan en los siguientes subtemas.

1. **Estructura Secuencial** Consiste en una secuencia finita de instrucciones elementales. Las instrucciones elementales pueden ser de dos tipos, asignaciones y utilización de recursos abstractos.
  - a) Asignación: Mediante esta operación se almacena en una variable (que básicamente identifica por un nombre una posición de memoria) el resultado de una operación aritmética o lógica.
  - b) Recursos abstractos: Consisten básicamente en una llamada a un procedimiento (un algoritmo al que se identifica por un nombre), al que se le pasan una serie de parámetros y devuelve otros.
2. **Estructura de Selección** Contiene fundamentalmente tres elementos: un predicado y una o dos estructuras secuenciales (con otras estructuras en su interior si hace falta). Un predicado es una sentencia que puede evaluarse como verdadera o falsa. Los predicados están formados tanto por expresiones booleanas que relacionan varios subpredicados como por expresiones relationales (mayor que, menor que, igual, diferente) que relacionan entre sí constantes y variables numéricas o alfanuméricas.

**3. Estructura Repetitiva** Está formada fundamentalmente por una estructura secuencial (con otras estructuras embebidas si es necesario) y un predicado. La estructura iterativa se ejecutará tantas veces como sea necesario hasta que el predicado se evalúe como verdadero (o como falso dependiendo del tipo). Los programas que utilizan sólo estas tres instrucciones de control básicas o sus variantes (como los bucles for, repeat o la instrucción condicional switch-case), pero no la instrucción goto, se llaman estructurados. La visión clásica de la programación estructurada se refiere al control de ejecución. La regla general es que las instrucciones se ejecuten sucesivamente una tras otra, pero diversas partes del programa se ejecutan o no dependiendo de que se cumpla alguna condición. Además, hay instrucciones (los bucles) que deben ejecutarse varias veces, ya sea en número fijo o hasta que se cumpla una condición determinada.

Si bien algunos pueden no estar de acuerdo con una definición exacta de SP (McCracken, 1973), existe consenso sobre sus dos características principales: (1) un programa estructurado es una disposición jerárquica de módulos altamente independientes, y (2) el flujo dentro del programa está controlado exclusivamente por solo tres formas de control, **selección, repetición y secuenciación**. Las estructuras de control básicas propuestas para implementarlas son ‘IF THEN ELSE’ para la selección, ‘DO WHILE’ para la repetición, y simplemente colocando las afirmaciones una tras otra para la secuenciación.

## 2.2. Programación Orientada a Objetos (POO)

La programación estructurada y la orientación a objetos permiten a los programadores escribir código sin necesidad de conocer los detalles de la implementación de funciones de una librería y objetos respectivamente. Su implementación permanece oculta a los programadores, liberándoles de los condicionantes que esto imponía. Lo realmente complicado de ambos paradigmas de programación es encontrar los modelos o patrones conceptuales correctos para el problema y no la implementación de los mismos. Realmente, el diseño adquiere la máxima importancia que no tenía en la programación tradicional.

La propuesta de la moderna programación estructurada y de la orientación a objetos consiste principalmente en intentar superar la complejidad inherente a los problemas del mundo real gracias a la abstracción de nuestro conocimiento sobre tales problemas y su encapsulación dentro de módulos o librerías y objetos respectivamente. Se utiliza el término encapsular porque los objetos y las librerías pueden verse como cápsulas.<sup>a</sup> modo de barrera conceptual que separa un conjunto de valores y operaciones, que poseen un substrato conceptual idéntico, del resto del sistema.

Así, tanto objetos como librerías tienen dos partes bien diferenciadas: la declaración y la implementación. La declaración o interfaz es de dominio público, conocido por las demás partes del programa. En la implementación se explicitan y codifican los elementos necesarios para responder a lo especificado en la declaración. Esta separación permite que cada objeto definido en nuestro sistema pueda ocultar su implementación y hacer pública su interfaz de comunicación.

La programación orientada a objetos (POO) surge de la necesidad de contar con lenguajes que pudieran implementar soluciones computables de una forma más parecida a la que utilizamos en la vida real, ya que nosotros concebimos nuestro mundo como un conjunto de cosas u objetos con propiedades o atributos que los definen como tal.

## 2.3. Bases de Datos Relacionales

Existen diversos modelos de bases de datos. El modelo de la base de datos define un paradigma de almacenamiento, estableciendo cómo se estructuran los datos y las relaciones entre estos. Las distintas operaciones sobre la base de datos (eliminación o sustitución de datos, lectura de datos, etc.) vienen condicionadas por esta estructura, y existen notables diferencias entre los principales modelos, cada uno de ellos con sus ventajas e inconvenientes particulares.

Las Bases de datos relacionales constituyen el modelo de bases de datos más utilizado en la actualidad. Solucionan los problemas asociados a las bases de datos jerárquicas y en red, utilizando para ello un esquema basado en tablas, que resulta a la vez sencillo de comprender y fácil de utilizar para el análisis y la consulta de los datos. Las tablas contienen un número dado de registros (equivalentes a las filas en la tabla), así como campos (columnas), lo que da lugar a una correcta estructuración y un acceso eficiente.

Paralelamente, el modelo relacional se sustenta en unos fundamentos matemáticos sólidos y sus ideas pueden expresarse mediante conceptos de la teoría de conjuntos, lo que posibilita un análisis formal del mismo. Además de las denominaciones habituales de tabla, fila y columna, existe una terminología específica empleada al referirse a las bases de datos relacionales. Así, en el modelo relacional los datos se organizan en tablas bidimensionales, cada una de ellas con información relativa a un determinada entidad.

La tabla en sí se conoce como relación, ya que recoge la relación existente entre sus elementos, y constituye así el eje central del modelo relacional. Dentro de la tabla, los datos están organizados a su vez en filas y columnas. Las columnas representan los distintos atributos asociados a la entidad, mientras que las filas conforman los distintos registros. Una fila se forma con un conjunto de n atributos, constituyendo una tupla.

El esquema de la relación está formado por los nombres de los atributos y un dominio asociado a estos, que delimita el rango de valores posibles para cada atributo. El dominio especifica el tipo de dato a contener en cada columna. Por ejemplo, si se recoge un nombre el atributo será de tipo alfanumérico, mientras que si el atributo es un conteo deberá ser de tipo entero. Además de los tipos habituales (fechas, cadenas de texto, valores reales). Puede emplearse también la denominación menos formal de número decimal o bien valor de coma flotante, esta última más común en el ámbito informático y referida a la forma de almacenamiento de este tipo de valores, valores enteros, etc.) pueden emplearse en ciertas bases de datos valores más complejos.



---

# CAPÍTULO 3

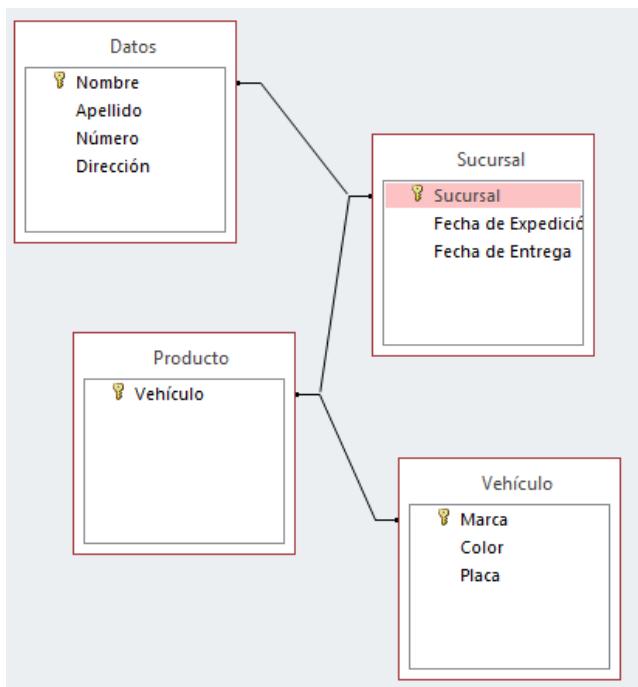
---

## MANEJO DE RDBMS

Un sistema de gestión de bases de datos relacionales (RDBMS en inglés) es un programa que te permite crear, actualizar y administrar una base de datos relacional. La mayoría de los RDBMS comerciales utilizan el lenguaje de consultas estructuradas (SQL) para acceder a la base de datos, aunque SQL fue inventado después del desarrollo del modelo relacional y no es necesario para su uso.

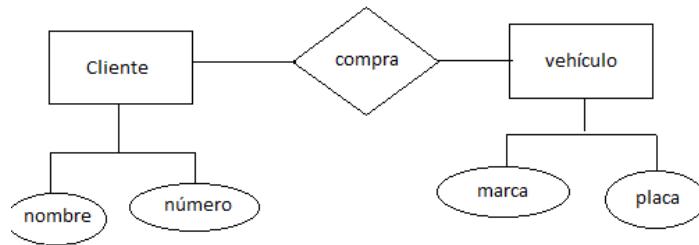
### 3.1. Diseños Relacionales

El diseño de una base de datos consiste en definir la estructura de los datos que debe tener un sistema de información determinado. Para ello se suelen seguir por regla general unas fases en el proceso de diseño, definiendo para ello el modelo conceptual, el lógico y el físico.



**Figura 3.1:** Ejemplo de Diseño Racional

**Modelo conceptual:** En ésta fase, se plasman las entidades y las relaciones que existirán entre ellas.



**Figura 3.2:** Ejemplo de Modelo Conceptual

**Modelo lógico:** Aquí, se tabula el diseño conceptual, así que se debe pensar en cómo normalizar las tablas para evitar duplicidad de información y ahorrar espacio de almacenamiento.

**Modelo físico:** En esta última fase ya debemos revisar a detalle los tipos de datos que utilizaremos, sus dominios (qué valores va a permitir), cuales índices debemos crear para optimizar las consultas, entre otros. Aquí ya escribimos nuestro SQL para plasmar todo nuestro diseño en el motor de bases de datos elegido.

```

CREATE TABLE clientes (
    id CLIENTE,
    Nombre VARCHAR(15) NOT NULL,
    Apellido VARCHAR(20) NOT NULL,
    Número INT,
    CONSTRAINT id_Cliente PRIMARY KEY(id)
);
    
```

**Figura 3.3:** Ejemplo de Modelo Físico

## 3.2. Diseños para Minería de Datos

La minería de datos es el conjunto de técnicas y herramientas aplicadas a procesos de acceso, presentación y extracción de información a partir de bancos de datos en forma automatizada. La importancia de la minería de datos radica en el estudio y aplicación de las técnicas que permitan una eficiencia que responda al crecimiento en el manejo de grandes volúmenes de datos.

1. Selección del conjunto de datos, tanto en lo que se refiere a las variables dependientes, como a las variables objetivo, como posiblemente al muestreo de los registros disponibles.
2. Análisis de las propiedades de los datos, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).
3. Transformación del conjunto de datos de entrada, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema.

4. Evaluar los resultados contrastándolos con un conjunto de datos previamente reservado para validar la generalidad del modelo.

Si el modelo final no superara esta evaluación el proceso se podría repetir desde el principio o, si el experto lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo válido.

### 3.3. Procesos de Extracción-Transformación de Carga (ETL)

Extraer, transformar y cargar (ETL) es un proceso de almacenamiento de datos que implica extraer datos de orígenes externos, transformarlos para satisfacer las necesidades de la empresa y finalmente, cargarlos en el depósito de datos. Es importante que se realicen bien estos pasos para así poder garantizar la veracidad, integridad y centralización de los datos.

1. Proceso de Extracción: Normalmente, hay una primera etapa de análisis donde se evalúan las necesidades concretas de la organización en cuanto a movimiento y transformación de datos. Antes de extraer los datos, se identifica de qué tipo son y en qué formato se encuentran los sistemas fuente y en función de las necesidades detectadas, se procede a la extracción en sí de dichos datos.
2. Proceso de Transformación: Existen tres modos distintos de extracción, se elige dependiendo la necesidad de la organización es decir: *Full Extract o extracción total, Incremental Extract o extracción incremental y Update Notification o notificación de actualizaciones*
3. Proceso de Carga: En esta fase, los datos procedentes de la fase anterior son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes.

Mes	Capital	Interés	Monto
Enero	3.000,00	450,00	3.450,00
Febrero	3.450,00	517,50	3.967,50
Marzo	3.967,50	595,13	4.562,63
Abril	4.562,63	684,39	5.247,02
Mayo	5.247,02	787,05	6.034,07
Junio	6.034,07	905,11	6.939,18
Julio	6.939,18	1.040,88	7.980,06
Agosto	7.980,06	1.197,01	9.177,07
Septiembre	9.177,07	1.376,56	10.553,63
Octubre	10.553,63	1.583,04	12.136,67
Noviembre	12.136,67	1.820,50	13.957,17
Diciembre	13.957,17	2.093,58	16.050,75

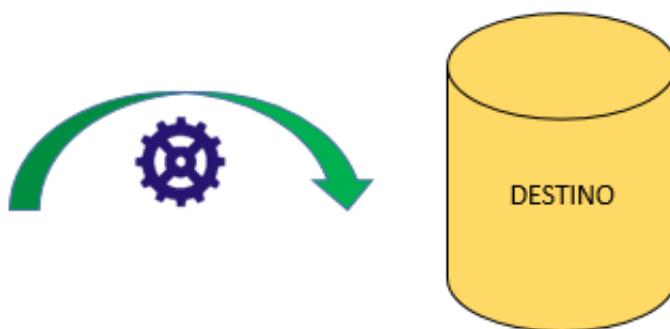


Figura 3.4: Proceso de Extracción de Carga



---

---

# CAPÍTULO 4

---

## INTELIGENCIA DE NEGOCIOS

### 4.1. Introducción

#### ¿Qué es Inteligencia de Negocios?

Según el Data Warehouse Institute, lo define como la combinación de tecnología, herramientas y procesos que permiten transformar datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados.

### 4.2. Indicadores clave de desempeño (KPI's)

Los KPI o indicadores clave de desempeño son mediciones financieras o no financieras que se utilizan para medir el grado de cumplimiento de los objetivos previamente establecidos, generalmente están contenido en el plan estratégico de la organización y reflejan su rendimiento en un periodo determinado; los KPIs se emplean en diversas áreas de una empresa entre las que se puede mencionar compras, ventas, logística y servicio y atención al cliente.

Los KPIs se utilizan para determinar el estado actual de un negocio y permiten definir una línea de acción futura; expresan las variables a tomar en cuenta para incrementar el desempeño favorable para conseguir el progreso, posicionamiento y ganancias significativas de la empresa. El KPI mide las condiciones para el rendimiento y el desempeño de manera clave.

Es importante considerar una serie de parámetros que deben fijarse para cada KPI que se desea crear, entre ellos:

1. Definición y Actividad que evalúa.
2. Forma de medirlo – ratio y Periodicidad.

#### 4.2.1. ¿Por qué los KPI definen el éxito?

Los KPI o indicadores clave del rendimiento o del desempeño son herramientas fundamentales del management y consultores del marketing, que son usadas para la ejecución de una estrategia empresarial y para visualizar cómo se proyectará en el futuro, desde el punto de vista de su rentabilidad.

Los KPI no miden resultados finales, sino cotejan aspectos claves del proyecto que incidirán de forma directa en su éxito y que permiten tomar acciones concretas en la empresa y generar condiciones más favorables en sus métodos.

Esto nos indica que no todas las métricas con las que se mide el desempeño deben ser tomadas como un KPI, ya que éstos son sólo los claves e indispensables para obtener el éxito. Están íntimamente relacionados con los factores claves del éxito, que permiten que las empresas se mantengan vigentes y además progresen de manera permanente.

#### 4.2.2. ¿Cómo utilizan los KPI las compañías garantizadas?

Las empresas exitosas cuentan con factores clave del éxito que las identifica por sus niveles de excelencia y que sus competidoras no pueden conseguir afinar en sus términos. Como la capacidad de renovarse de manera oportuna, con nuevos y fabulosos productos o el poder de adaptarse a las variaciones del mercado.

Un buen analista empresarial logra descubrir y conocer al detalle los KPI de los cuales depende el éxito de cada empresa. Es decir, éstos son aspectos intrínsecos o exclusivos de cada proceso productor y están allí para ser aprovechados de forma correcta para asegurar el éxito.

Se amerita que precisen cuáles son los KPI claves en cada proceso para controlar cada estrategia, por lo que es conveniente que cada empresa consiga determinar cuáles son sus indicadores claves más oportunos y trascendentales para vigilar su comportamiento y asegurar que se están cumpliendo en el proceso y que aseguran el rendimiento óptimo.

Uno de los aspectos en que se utilizan mayormente los indicadores claves es el marketing digital, los KPIs de marketing permiten a las empresas medir de manera eficaz el desarrollo de sus campañas publicitarias para contar con un sistema de control y seguimiento de las acciones lanzadas y del trabajo realizado, lo que les permite tener una mayor comprensión del error y del éxito de las mismas.

#### 4.2.3. KPI's de marketing

Los Indicadores Clave de Rendimiento o KPIs asociados al marketing digital son esenciales para saber cómo marcha una campaña publicitaria y si las metas establecidas se están cumpliendo en los porcentajes esperados, para ello se deben elegir indicadores que resulten realmente útiles para medir la efectividad de las acciones emprendidas.

## 4.3. Diseño de Tableros

Conocido también como Cuadro de Mando Integral (CMI) o tablero de comando o balanced scorecard. El tablero de control o Balance Scorecard (BSC) es una metodología gerencial que sirve como herramienta para la planeación y administración estratégica de las empresas. Es una aplicación de sistemas de autocontrol y mejora continua.

Lo podemos definir como una estructura de control de la administración y operación general de la empresa, cuya fortaleza radica en su filosofía de mejora continua y en el trabajo en equipo basado en una visión estratégica unificada. Al implantar el tablero de control se utilizan criterios de medición e indicadores para controlar la eficiencia y eficacia en el cumplimiento de la visión, misión y objetivos de la empresa.

Los objetivos principales del tablero de control son:

1. Medir los avances y cumplimiento de la visión, la misión, los valores, los objetivos y las estrategias de la empresa.
2. Alinear los indicadores y las metas de la dirección con la cadena de valor de la empresa y los indicadores y metas de las áreas.
3. Integrar el plan estratégico de la empresa con los planes operativos de las áreas.



---

# CAPÍTULO 5

---

## INTELIGENCIA ANALITICA

La inteligencia analítica es una herramienta que va a permitir desarrollar estrategias de negocios para así poder optimizar resultados críticos y no triviales en las diferentes áreas de una empresa. Claro estas enfocándose principalmente al crecimiento e implementación para esta.

En esta se ve mucho más aplicado el hecho de conocer a los clientes que esta empresa maneje ya que solo de esta forma va a ser posible hacer crecer la empresa, es decir conocer los nuevos intereses o necesidades que aún tenga esta clientela. Y de esta forma además de crear un crecimiento para la empresa también se estaría innovando y como debe de ser una empresa nunca debe de quedarse estancada siempre estamos en un constante cambio y más aún en la tecnología.

Ya que esta disciplina no se dedica únicamente a monitorear o saber cómo van los aspectos más importantes a analizar, esta tiene también como objetivo predecir cuestiones financieras y analizar lo que esté pasando, encontrar una razón a todo. Y con esta disciplina de inteligencia analítica hacemos un completo uso de la tecnología para así tener la capacidad de explotar datos de los clientes que ya se manejan y conocerlos mejor aún. Aunque con ayuda de la inteligencia analítica no solo va a ser posible conocer mejor a los clientes o hacer un análisis sobre ellos también va a ser posible analizar otro tipo de cuestiones en una empresa como puede llegar a ser el costo de producción.

En otras palabras, esta disciplina nos llevaría a un crecimiento económico y así proveer de una mejor microeconomía y esta comienza a usarse en más y más empresas el país estaría generando un mucho mayor crecimiento económico.

### 5.1. Metodología del Modelado

Es una herramienta usada para una especificación formal de Modelos orientados a objetos. Esta lleva una fase de identificación y conocimiento de los componentes del sistema que serán considerados y así poder proponer una arquitectura del sistema. Estos componentes y sus interrelaciones son especificados e implementados en las últimas fases de la metodología de la explotación de datos.

Aunque también debe saberse que para poder hacer uso de esta herramienta los datos deben de tener una estructura específica para que ya con estos sea posible hacer uso de la metodología de modelado.

Existen dos tipos de entrenamiento de modelado el aprendizaje supervisado y el no supervisado. En el primero se cuenta ya con un conjunto de ejemplos (conjunto de entrenamiento) clasificados que es a los que se les va a intentar asignar una clasificación según el grupo de ejemplos con el que se cuente, para el segundo no se cuenta con el grupo de ejemplos, en este únicamente a partir de los datos introducidos se intenta agrupar según la similaridad de estos.

### 5.1.1. Aprendizaje Automático

El Aprendizaje Automático (AA, o Machine Learning, por su nombre en inglés) es la rama de la Inteligencia Artificial que tiene como objetivo desarrollar técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear algoritmos capaces de generalizar comportamientos y reconocer patrones a partir de una información suministrada en forma de ejemplos. Es, por lo tanto, un proceso de inducción del conocimiento, es decir, un método que permite obtener por generalización un enunciado general a partir de enunciados que describen casos particulares.

### 5.1.2. Árboles de Desición

Es un método analítico que a través de una representación esquemática de las alternativas disponible facilita la toma de mejores decisiones, especialmente cuando existen riesgos, costos, beneficios y múltiples opciones. El nombre se deriva de la apariencia del modelo parecido a un árbol y su uso es amplio en el ámbito de la toma de decisiones bajo incertidumbre (Teoría de Decisiones).

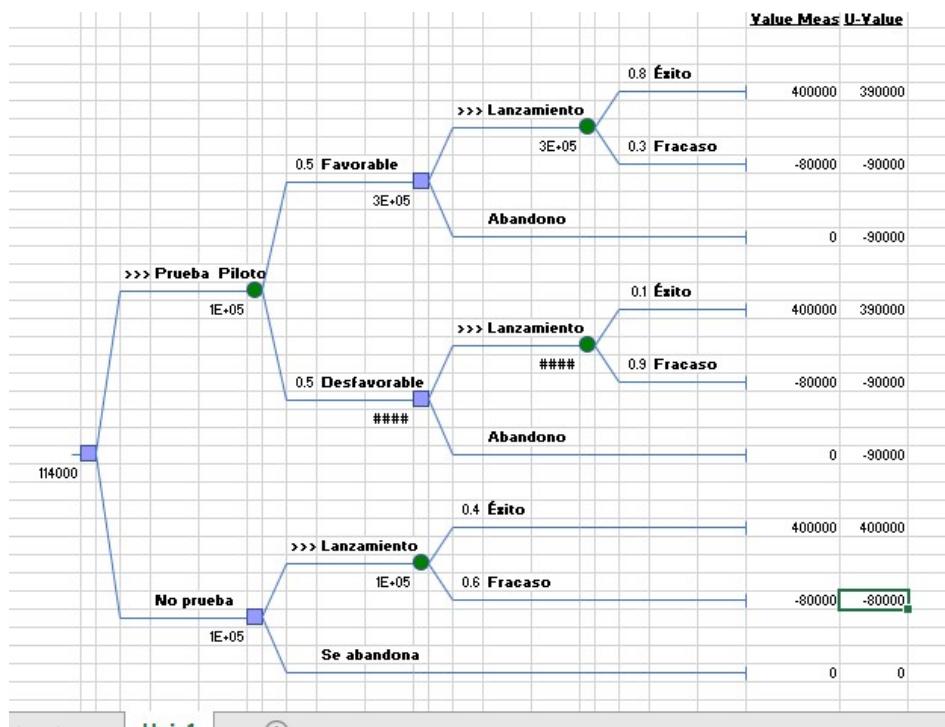


Figura 5.1: Ejemplo de Árbol de desición

### 5.1.3. Regresión

En estadística, el análisis de la regresión es un proceso estadístico para estimar las relaciones entre variables. Incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes (o predictoras). Más específicamente, el análisis de regresión ayuda a entender cómo el valor de la variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las otras variables independientes fijas.

El análisis de regresión es ampliamente utilizado para la predicción y previsión, donde su uso tiene superposición sustancial en el campo de aprendizaje automático. El análisis de regresión se utiliza también para comprender cuales de las variables independientes están relacionadas con la variable dependiente, y explorar las formas de estas relaciones. En circunstancias limitadas, el análisis de regresión puede utilizarse para inferir relaciones causales entre las variables independientes y dependientes. Sin embargo, esto puede llevar a ilusiones o relaciones falsas, por lo que se recomienda precaución,<sup>1</sup> por ejemplo, la correlación no implica causalidad.

### 5.1.4. Redes Neuronales

Las redes neuronales son un campo muy importante dentro de la Inteligencia Artificial. Inspirándose en el comportamiento conocido del cerebro humano (principalmente el referido a las neuronas y sus conexiones), trata de crear modelos artificiales que solucionen problemas difíciles de resolver mediante técnicas algorítmicas convencionales.

#### Modelo neuronal de McCulloch-Pitts

El primer modelo matemático de una neurona artificial, creado con el fin de llevar a cabo tareas simples, fue presentado en el año 1943 en un trabajo conjunto entre el psiquiatra y neuroanatomista Warren McCulloch y el matemático Walter Pitts.

### 5.1.5. Máquinas Vector Soporte

Las Máquinas de Vector Soporte (MVS) es un área prometedora del aprendizaje maquinal desarrollada inicialmente por Vapnik, para construir clasificadores. Se funda en la teoría estadística del aprendizaje, o teoría de VC. Esta teoría permite escoger un clasificador que minimiza una cota superior sobre el riesgo (o error de prueba), y proporciona una buena medida para obtener clasificadores que generalizan bien sobre datos no previamente vistos.

### 5.1.6. K vecinos más cercanos

El algoritmo clasifica cada dato nuevo en el grupo que corresponda, según tenga k vecinos más cerca de un grupo o de otro. Es decir, calcula la distancia del elemento nuevo a cada uno de los existentes, y ordena dichas distancias de menor a mayor para ir seleccionando el grupo al que pertenecer. Este grupo será, por tanto, el de mayor frecuencia con menores distancias.

### 5.1.7. Clasificador ingenuo de Bayes

En términos simples, un clasificador de Bayes ingenuo asume que la presencia o ausencia de una característica particular no está relacionada con la presencia o ausencia de cualquier otra característica, dada la clase variable. Por ejemplo, una fruta puede ser considerada como una manzana si es roja, redonda y de alrededor de 7 cm de diámetro. Un clasificador de Bayes ingenuo considera que cada una de estas características contribuye de manera independiente a la probabilidad de que esta fruta sea una manzana, independientemente de la presencia o ausencia de las otras características.

Para otros modelos de probabilidad, los clasificadores de Bayes ingenuo se pueden entrenar de manera muy eficiente en un entorno de aprendizaje supervisado. En muchas aplicaciones prácticas, la estimación de parámetros para los modelos Bayes ingenuo utiliza el método de máxima verosimilitud, en otras palabras, se puede trabajar con el modelo ingenuo de Bayes sin aceptar probabilidad bayesiana o cualquiera de los métodos bayesianos

Una ventaja del clasificador de Bayes ingenuo es que solo se requiere una pequeña cantidad de datos de entrenamiento para estimar los parámetros (las medias y las varianzas de las variables) necesarias para la clasificación. Como las variables independientes se asumen, solo es necesario determinar las varianzas de las variables de cada clase y no toda la matriz de covarianza.

## 5.2. Ensamble de Modelos

Cuando las personas tienen que tomar decisiones difíciles, normalmente toman en cuenta la opinión de varios expertos, buscando mejorar sus decisiones, en la ciencia de datos, cada modelo inducido se puede ver como un “experto”, por lo que uno puede pensar en que se puedan producir predicciones más confiables si se combinan las predicciones de varios modelos, para poder mejorar una toma de decisiones, que en particular este proceso es llamado “**el ensamble de modelos**”, el cual busca incrementar la exactitud a partir de la combinación de las predicciones de múltiples modelos.

### 5.2.1. Bosques Aleatorios

Es una combinación de árboles de decisión tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de los mismos. Se promediar muchos modelos ruidosos, pero aproximadamente imparciales, lo que implicaría reducir la variación puesto ya que ellos pueden registrar estructuras de interacción compleja en los datos.

#### Ventajas:

1. Ser uno de los algoritmos de aprendizaje más certeros que hay disponible.
2. Trabajan eficientemente con grandes datos.
3. Manejar cientos de variables de entrada sin excluir ninguna.

### 5.2.2. AdaBoost(Adaptative Boosting), Impulso Adaptativo

Hace referencia a un tipo de algoritmos cuya finalidad es encontrar una hipótesis fuerte a partir de utilizar hipótesis simples y débiles, es decir, busca crear un clasificador fuerte, cuya base sea la combinación lineal de clasificadores débiles, en el cual propone entrenar de forma iterativa, de modo que cada nuevo clasificador se enfoque en los datos que fueron erróneamente clasificados por su predecesor, de esta forma el algoritmo se adapta y logra obtener mejores resultados.

### 5.2.3. Impulso de Árbol Gradiente

El impulso de árboles gradiente es una técnica de aprendizaje automático, que se utiliza para problemas en el campo del aprendizaje para clasificar o regresión, La idea principal es diseñar un modelo de predicción en forma de un conjunto de modelos de predicciones débiles, (usualmente usados en árboles de decisión).

El impulso de árboles gradiente construye el modelo de manera escalonada como lo hacen otros métodos de impulso ( como por ejemplo impulso adaptativo), y los generaliza al permitir la optimización de una función de pérdida diferenciable arbitraria tomando en cuenta los residuos que se obtienen en las estimaciones los cuales son implícitamente agregados a una función de costo adecuada.

	Regresión Lineal	Regresión Logística	Árbol de decisión	Árbol Aleatorio	Árbol Gradiente Boost	Redes Neuronales
DESCRIPCIÓN	Estudia la relación entre variables	La adaptación de la regresión lineal a problemas de clasificación a los problemas de adaptación	Gráfica que ramifica todos los resultados posibles de una decisión	Toma el promedio de muchos árboles de decisión para generar uno más potente	Técnica de aprendizaje automático utilizado para el análisis de regresión y problemas estadísticos	Permiten extraer información útil con los datos disponibles
VENTAJAS	Se puede hacer una predicción del comportamiento de alguna variable en un determinado punto o momento	Fácil de entender	Fácil de entender y usar	Arroja modelos de muy alta calidad y es rápido de entrenar	Es de alto rendimiento	Puede manejar tareas extremadamente complejas
DESVENTAJAS	Demasiado simple para capturar relaciones complejas entre las variables	Tendencia a que el modelo se "adapte"	No es un método que se pueda usar en datos complejos	Puede ser lento para generar predicciones	Un pequeño cambio en el entrenamiento puede generar cambios radicales en el modelo.	Muy muy lento de entrenar debido a que tiene muchas capas

Figura 5.2: Diversos Modelos y Algoritmos



---

# CAPÍTULO 6

---

## DESARROLLO DE LA APLICACIÓN DE NEGOCIO

**ACA-BUS** surge con el propósito de brindar un servicio de calidad y sobre todo de seguridad a la comunidad de la Facultad de Estudios Superiores Acatlán, dicho esto, por estadísticas se sabe que en los últimos años un porcentaje alto de la comunidad a sufrido asaltos en diversos niveles, agresiones físicas e incluso verbales. Con lo anterior **ACA-BUS** busca una manera de resolver dicha problemática por lo que pone al servicio de la población de Acatlán transportes con rutas estratégicas y seguras para terminar con la problemática ya antes mencionada.

### 6.1. MISIÓN

Proporcionar a todos nuestros clientes un servicio de transporte seguro, oportuno y cómodo, a un precio competitivo y con una auténtica atención personalizada.

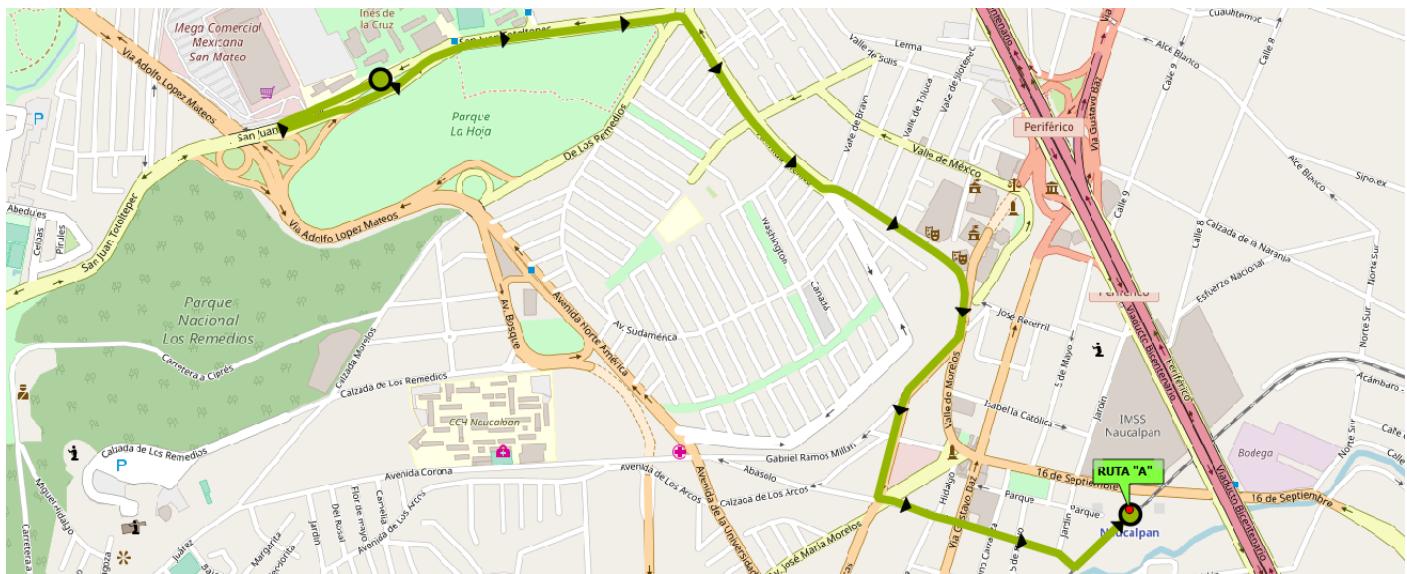
### 6.2. VISIÓN

Nuestra visión es consolidarnos como la empresa líder en el transporte Universitario gracias a nuestra calidad, prestigio, mejoramiento continuo y respuesta en el servicio en la Ciudad de México y Estado de México.

### 6.3. MODO DE OPERACIÓN

**ACA-BUS** estará laborando en colaboración con la Empresa **UTEP**, empresa fundada en 1934, quién desde 2001 cuenta con un Centro de Capacitación y Adiestramiento de conductores del servicio de autotransportes federal privado, acreditado por la Secretaría de Comunicaciones y Transportes. El objetivo de trabajar con **UTEP** es para proporcionar un servicio más completo para nuestra comunidad de Actlán.

**ACA-BUS**, realizó un estudio, con el cuál pondrá al servicio de la comunidad de Acatlán 7 rutas que se describen de forma detallada a continuación, esperando expandir esas rutas en un futuro no muy lejano.



**Figura 6.1:** RUTA A:NAUCALPAN DE JUÁREZ

**RUTA .<sup>A</sup>:** NAUCALPAN DE JUÁREZ

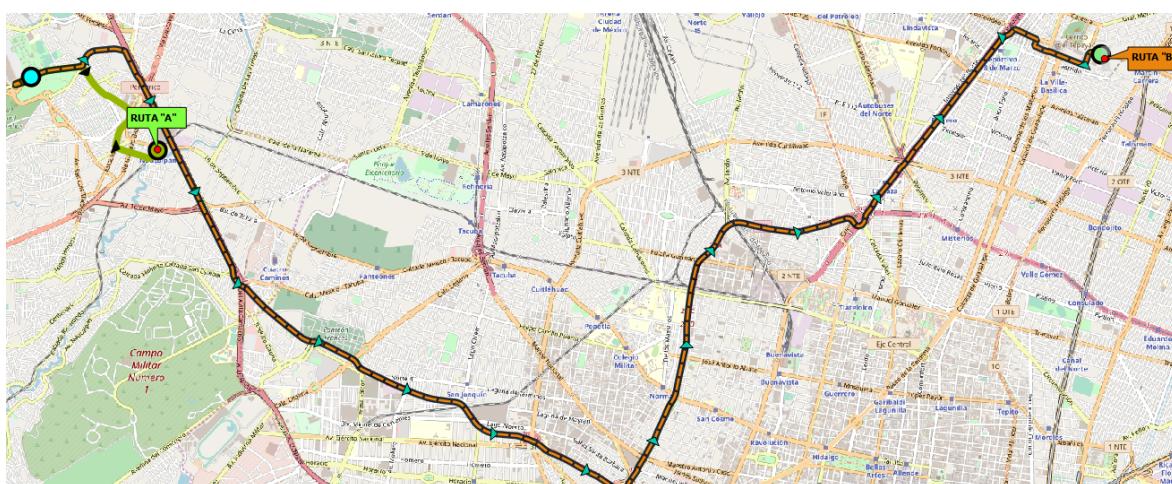
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 2.63924KM

**COSTO:**\$11.

**DESTINO:** AVENIDA VIADUCTO BICENTENARIO (PERIFERICO)

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.2:** RUTA B: GUSTAVO A. MADERO

**RUTA "B":** GUSTAVO A. MADERO

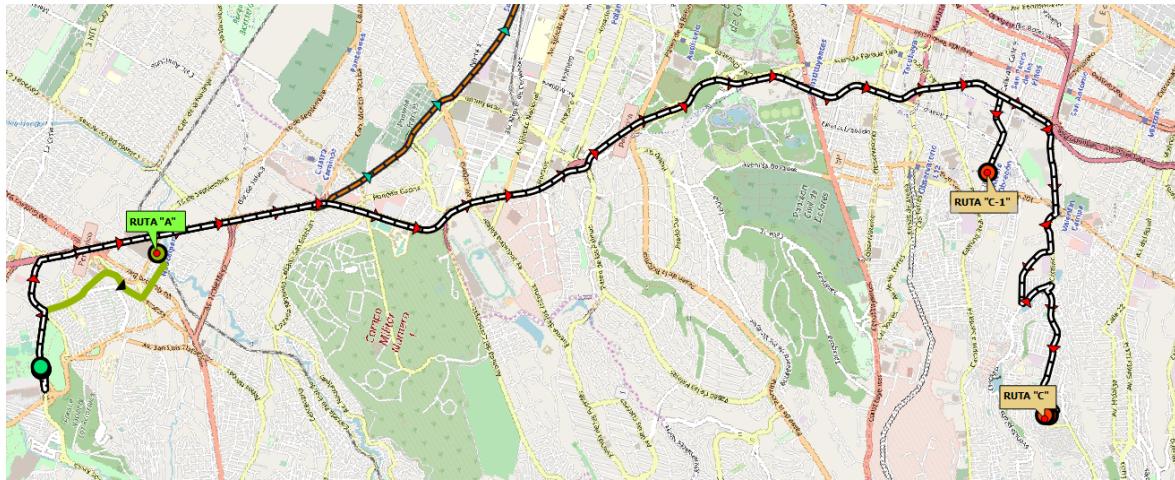
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 21.0025KM

**COSTO:** \$20,00

**DESTINO:** AVENIDA GUSTAVO A. MADERO-ESQUINA CALLE PRIMERO DE MAYO

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.3: RUTA C:ÁLVARO OBREGÓN**

#### RUTA C": ÁLVARO OBREGÓN

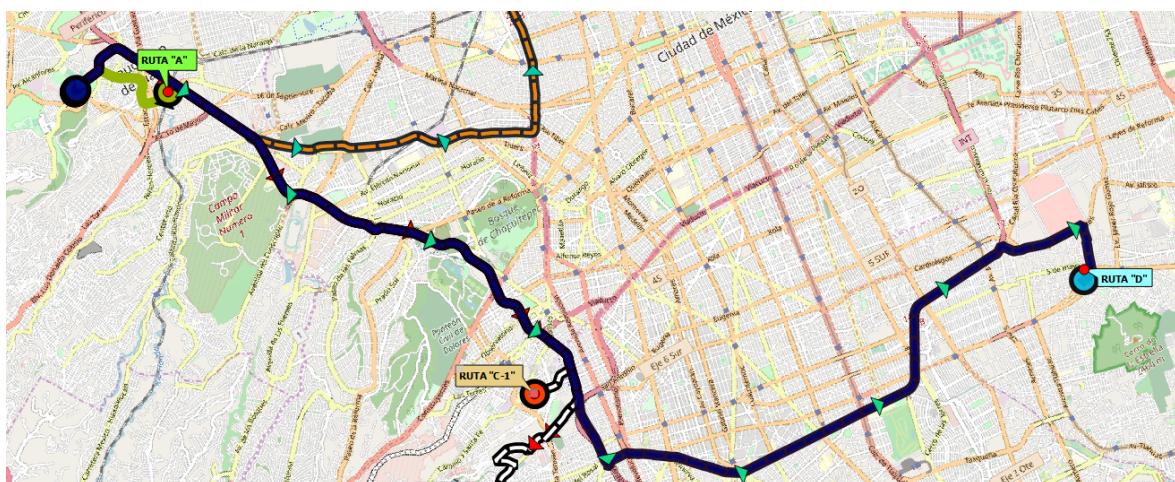
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 8.57378, 11.0793KM

**COSTO:** \$15,5

**DESTINO:** AVENIDA ÁLVARO OBREGÓN-ESQUINA ÁLVARO OBREGÓN-1

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.4: RUTA D:IZTAPALAPA**

#### RUTA "D": IZTAPALAPA

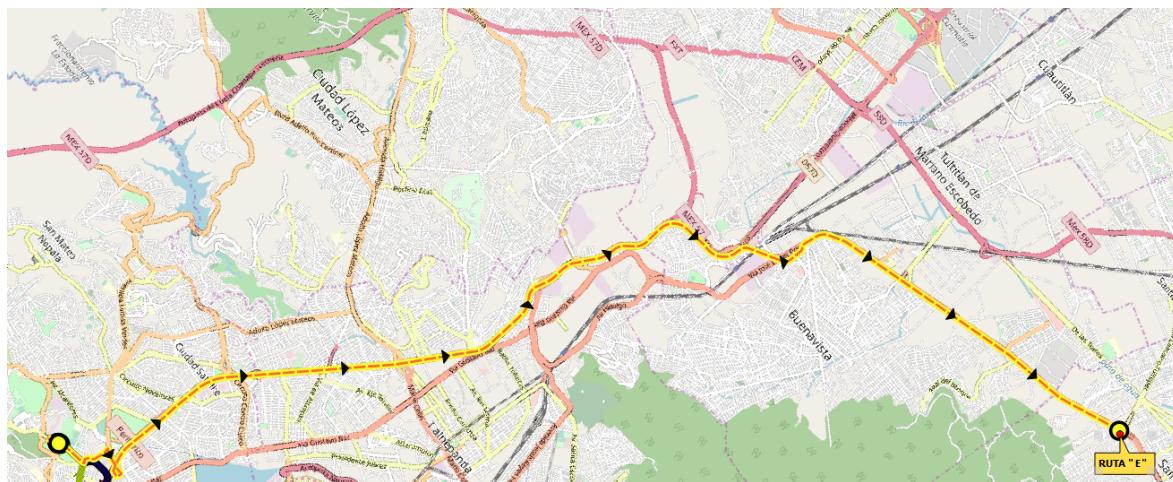
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 11.0417km

**COSTO:**\$15,5

**DESTINO:** AVENIDA SAN LORENZO

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.5: RUTA E:COACALCO DE BERRIOZABAL**

**RUTA .E:** COACALCO DE BERRIOZABAL

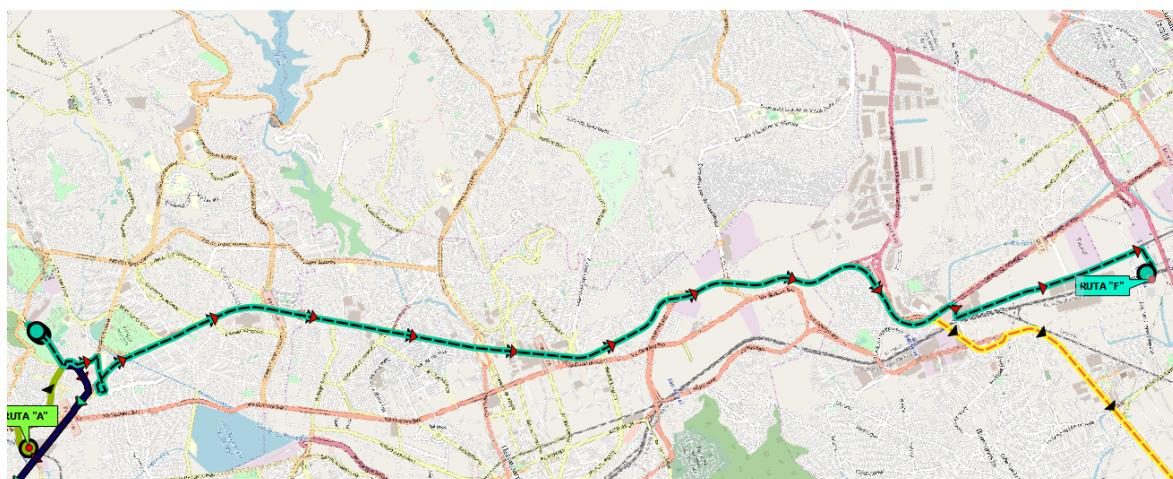
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 11.0417km

**COSTO:**\$15,5

**DESTINO:** AVENIDA JOSÉ LOPEZ PORTILLO

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.6: RUTA F:TULTITLÁN**

**RUTA "F":** TULTITLÁN

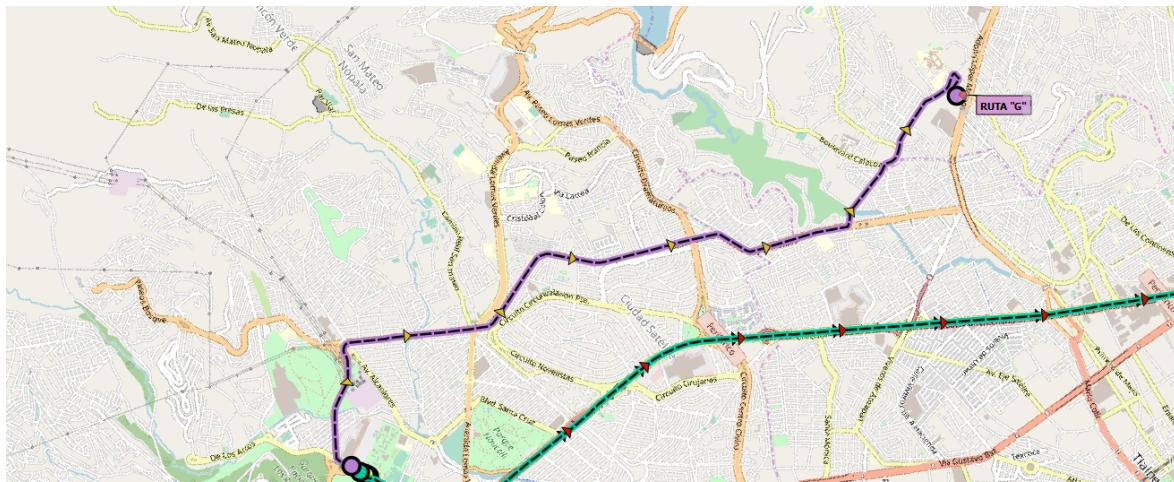
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 11.0417km

**COSTO:**\$19,0

**DESTINO:** AVENIDA JOSÉ LOPEZ PORTILLO

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.7: RUTA G:ATIZAPAN DE ZARAGOZA**

**RUTA "G": ATIZAPAN DE ZARAGOZA**

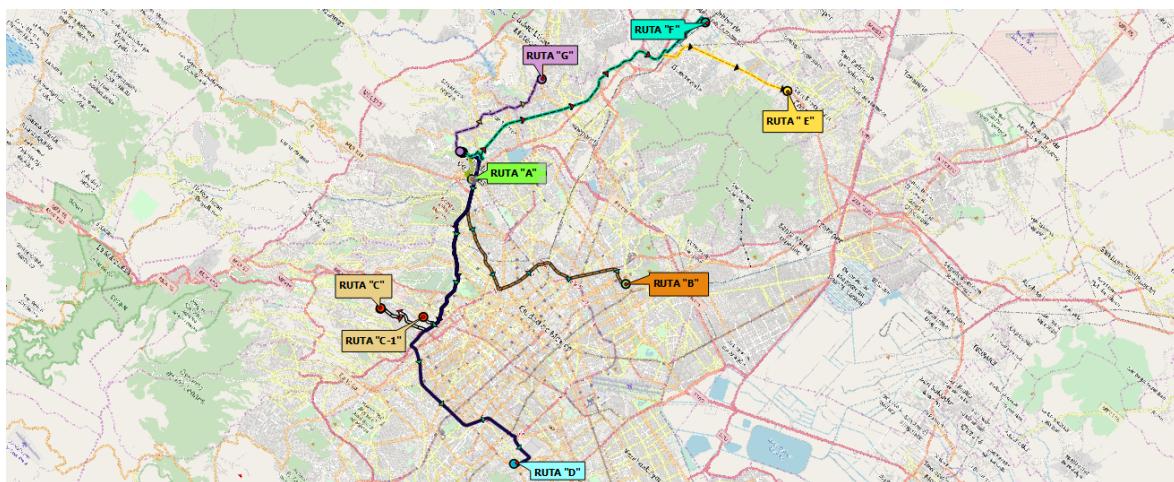
**SALIDA:** FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**TRAYECTORIA:** 1.0223km

**COSTO:\$10**

**DESTINO:** AVENIDA PDTE. ADOLFO LÓPEZ MAETOS

**HORARIOS:** 8:00PM - 10:00 PM



**Figura 6.8: RUTAS GENERALES**

**ACA-BUS** brindará rutas tanto para alumnos que residan en el Estado de México y para alumnos que residan en las diversas Alcaldías de la Ciudad de México. Las rutas están diseñadas con el objetivo de reducir el mayor tiempo posible (km). proporcionando una ubicación en avenidas principales a costos accesibles.

## 6.4. SEGURIDAD ESTUDIANTIL

**ACA-BUS**, buscará un convenio con las autoridades de la Facultad de Estudios Superiores Acatlán para llevar un mejor control del alumnado, el objetivo será tener un control de estudiantes, pues el servicio será de uso **EXCLUSIVO** para la comunidad, lo anterior se sustenta con base en brindar

mayor seguridad a los estudiantes.

**ACA-BUS** brindará tarjetas (credenciales de transporte) a los alumnos que decidan hacer uso de nuestro servicio, dichas credenciales serán las vías de pago para poder acceder a los autobuses, lo anterior con el fin de tener el control de que sólo sean estudiantes quienes hagan uso del servicio, mismas que podrán ser adquiridas en pequeños establecimientos que se pondrán dentro de la Universidad y dónde se podrán recargar.

**ACA-BUS**, cuenta con una aplicación móvil, dónde los estudiantes podrán estar al tanto de los siguientes aspectos:AVISOS, RUTAS,RECARGAR LA TARJETA,REGISTRO DE USUARIO, QUEJAS Y SUGERENCIAS.

Las siguiente imagen representan la estructura de la aplicación:

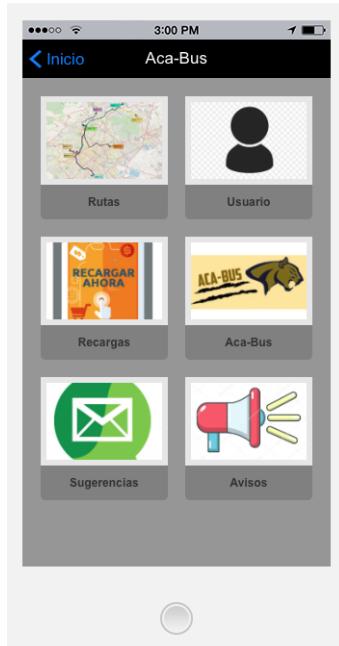


Figura 6.9: APP-ACABUS

## 6.5. IMPLEMENTACIÓN DE MACHINE LEARNING

**ACA-BUS** implemento un modelo de Regresión Línea para predecir si la comunidad Universitaria de Acatlán utilizaría nuestro servicio o no, dando resultados satisfactorios.

---

# CAPÍTULO 7

---

## ANÁLISIS EXPLORATORIO

En esta parte se muestran los histogramas, las frecuencias, correlaciones, outliers y tratamiento de missings obtenidos.

```
In [150]: import pandas as pd
import matplotlib.pyplot as plt
import csv

import re
import unicodedata
import seaborn as sns
%matplotlib inline
import numpy as np

In [151]: df=pd.read_csv('BaseTemas.csv',encoding='utf8')

In [152]: df.shape

Out[152]: (790, 38)

In [153]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 790 entries, 0 to 789
Data columns (total 38 columns):
SEXO           790 non-null int64
EDAD          790 non-null int64
CARRERA        790 non-null object
DM             790 non-null object
1_TRANSPI      790 non-null object
```

**Figura 7.1:** Limpieza Datos Parte - 1

Comenzamos con una población de 860 datos recolectados, de los cuáles, nos quedamos con la cantidad de 790 datos para trabajar.

```
dtypes: float64(2), int64(18), object(18)
memory usage: 234.6+ KB

In [154]: df.describe()

Out[154]:
   SEXO      EDAD      3_DIAS_UNIV      4_TRANS_PUBLIC      5_ASALTO      6_NUM_VECES      9_REND_ACADEMICO      10_AUTO      11_US/
count  790.000000  790.000000  790.000000  790.000000  790.000000  790.000000  790.000000  790.000000  790.000000
mean   0.418987  20.373418  5.064557  0.982278  0.313924  0.741772  0.637975  0.162025
std    0.493706  3.105423  0.966949  0.132021  0.464380  1.711558  0.480891  0.368707
min    0.000000  15.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
25%    0.000000  18.000000  5.000000  1.000000  0.000000  0.000000  0.000000  0.000000  0.000000
50%    0.000000  20.000000  5.000000  1.000000  0.000000  0.000000  1.000000  0.000000  0.000000
75%    1.000000  22.000000  6.000000  1.000000  1.000000  2.000000  1.000000  0.000000  0.000000
max    1.000000  49.000000  6.000000  1.000000  1.000000  25.000000  1.000000  1.000000  1.000000

In [155]: var_cont=list(df.describe().columns)
var_disc = [v for v in df.columns if v not in var_cont]

In [156]: print(len(var_disc)) #Número de variables discretas
print(len(var_cont)) #Número de variables continuas
```

**Figura 7.2:** Limpieza Datos Parte - 2

Lo primero que hicimos fue dejar a las variables que se prestaban en formato binario, los Si - 1, No - 0. Para trabajar de una mejor manera.

```

Out[159]: <bound method DataFrame.info of      SEXO EDAD          CARRERA \\
0   1   21          Actuaría
1   0   20          Actuaría
2   1   18          Actuaría
3   1   23          Actuaría
4   1   20          Actuaría
5   0   22          Actuaría
6   1   20          Actuaría
7   1   18          Actuaría
8   1   21          Actuaría
9   1   22          Actuaría
10  0   22         Actuaría
11  0   18         Actuaría
12  0   22         Actuaría
13  0   22         Actuaría
14  1   21         Actuaría
15  1   21         Actuaría
16  1   21         Actuaría
17  0   22         Actuaría
18  0   22

In [160]: def clean_text(text):
    try:
        res = text.decode('utf-8')
    except:
        res = text
    res = unicodedata.normalize('NFD', res).encode('ascii', 'ignore').lower()
    res = re.sub("[^a-zA-Z0-9 ]","", res.decode("utf-8"), flags=re.UNICODE)
    res = u' '.join(res.lower().split())
    return res

```

**Figura 7.3:** Limpieza Datos Parte - 3

Nos dimos cuenta que teníamos una población mayor de la carrera de Actuaría y de la Delegación de Álvaro Obregón y Municipio de Naucalpán de Juárez.

```

vard = [u'CARRERA',u'DM',u'L_TRANS',u'PROBL_TRANS',u'LLEG_TARDE_UNIV',u'HUMOR',u'RAZONES',u'RAZON_TRANS',
varc = [u'SEXO',u'EDAD',u'TIEM_DEMORA',u'DIAS_UNIV',u'TRANS_PUBLIC',u'ASALTO',u'NUM_VECES',u'REND_ACADEMIC']

varc2=[c %d%i for i in range(len(varc))]
vard2=[d %d%i for i in range(len(vard))]

df.rename(columns=dict(zip(varc,varc2)),inplace=True)
df.rename(columns=dict(zip(vard,vard2)),inplace=True)

df = df[[v for v in df.columns if (v[:2] in ['c_','d_']) or (v==u'CARRERA')]]

df.head()

   c_0  c_1  d_0  d_1       d_2  c_2  c_3  c_4  c_5  c_6 ...  d_5  c_17  c_18  d_12  c_19  c_20  d_13  c_21
0   1   21 Actuaría Acambay  Microbus, 04:00hrs  5   1   0   0 ... Fastidioso  1   1   UBER   1   1   Seguridad,
                                         Bicicleta, Metro
                                         Comodidad, Ahorro de hasta 1 hora e...
                                         Responsabilidad...
1   0   20 Actuaría Acolman  Microbús 01:00hrs  6   1   0   0 ... Fastidioso  1   1   UBER   1   1   Seguridad,
                                         Comodidad, Puntualidad,
                                         Responsabilidad...
2   1   18 Actuaría Acolman  Automóvil propio, 01:00hrs  5   1   1   1 ... Fastidioso  1   1   UBER   1   0   Seguridad,
                                         Microbús A pie
                                         Puntualidad, Responsabilidad, Tran...

```

**Figura 7.4:** Limpieza Datos Parte - 4

Tuvimos que ir filtrando datos, es decir, corregimos el formato hora, le dimos el mismo tipo de letra para que fuesemos teniendo unas frecuencias más certeras.

```
: df = df.loc[df['c_0']!=df['c_0'].max()]

: df[vard2].describe()

d_0 d_1 d_2 d_3 d_4 d_5 d_6 d_7 d_8 d_9 d_10 d_11 d_12 d_13
count 459 459 459 459 459 459 459 459 459 459 459 459 459 459
unique 17 46 52 26 3 5 18 27 5 5 5 31 4
top Actuaría Naucalpan de Juárez Microbús Saturación de transporte Si, pocas veces Fastidiado NINGUNA No hay otro medio Todos los días Regular Regular Limpieza, Tiempo de Espera, Amabilidad, Condic... UBER Seguridad, Comodidad, Puntualidad, Limpieza
freq 227 97 87 110 269 304 420 153 334 279 302 107 270

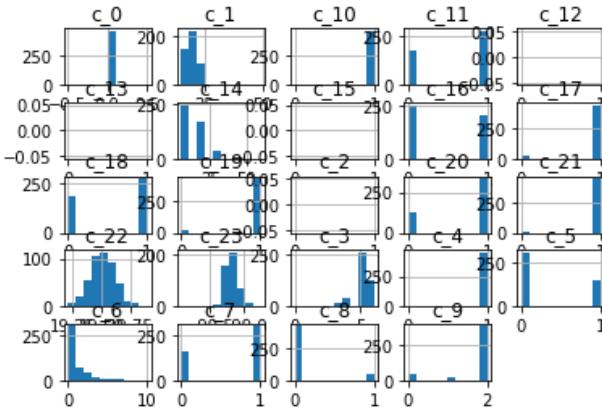
: vard2 = df[vard2]

: vard2.head(10)

d_0 d_1 d_2 d_3 d_4 d_5 d_6 d_7 d_8 d_9 d_10 d_11 d_12
1 Actuaría Acolman Microbús Acoso, Saturación de transporte Si, pocas veces Fastidiado NINGUNA No hay otro medio Todos los días Pésimo Malo Limpieza, Tiempo de Espera, Amabilidad, Condic... UBER p R
2 Acolman Actuaría Microbús Acoso, Saturación de transporte Si, pocas veces Fastidiado NINGUNA No hay otro medio Todos los días Pésimo Malo Limpieza, Tiempo de Espera, Amabilidad, Condic... UBER p R
```

**Figura 7.5:** Limpieza Datos Parte - 5

De igual manera, hicimos describir, para ver el comportamiento de nuestros datos, resaltando que la mayor de la población ya había sufrido algún tipo de asalto durante la trayectoria escolar.



**Figura 7.6:** Limpieza Datos Parte - 6

Realizamos el siguiente histograma por tipo de campo, y notamos como persistia cambios muy notorios en unas variables que en otras.

```
In [60]: X = df[[u'EDAD',u'3 DIAS_UNIV',u'6_NUM_VECES',u'9_REND_ACADEMICO',u'10_AUTO',u'22_MIN_ESPERA']].copy()

In [61]: sc = MinMaxScaler()

In [62]: sc.fit(X)

Out[62]: MinMaxScaler(copy=True, feature_range=(0, 1))

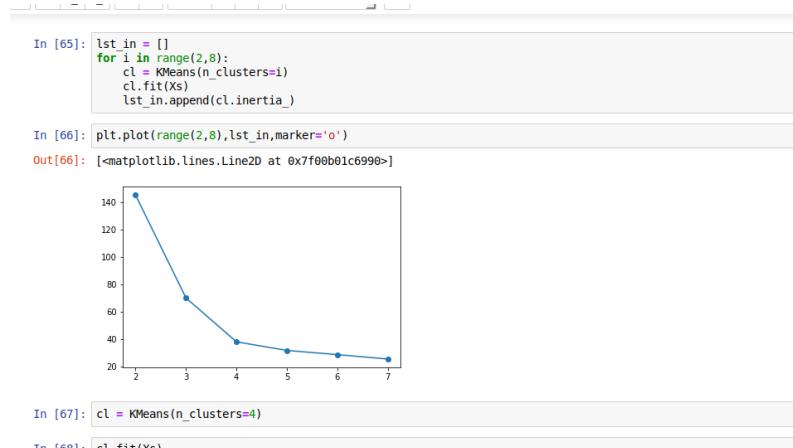
In [63]: Xs = pd.DataFrame(sc.transform(X),columns=X.columns)

In [64]: Xs.describe()

Out[64]:
   EDAD  3_DIAS_UNIV  6_NUM_VECES  9_REND_ACADEMICO  10_AUTO  22_MIN_ESPERA
Count    790.000000      790.000000      790.000000      790.000000      790.000000
mean     0.158042      0.844093      0.029671      0.637975      0.162025      0.102755
std      0.091336      0.161158      0.068462      0.480891      0.368707      0.102408
min      0.000000      0.000000      0.000000      0.000000      0.000000      0.000000
25%      0.088235      0.833333      0.000000      0.000000      0.000000      0.000000
50%      0.147059      0.833333      0.000000      1.000000      0.000000      0.058824
75%      0.205882      1.000000      0.040000      1.000000      0.000000      0.176471
max      1.000000      1.000000      1.000000      1.000000      1.000000      1.000000
```

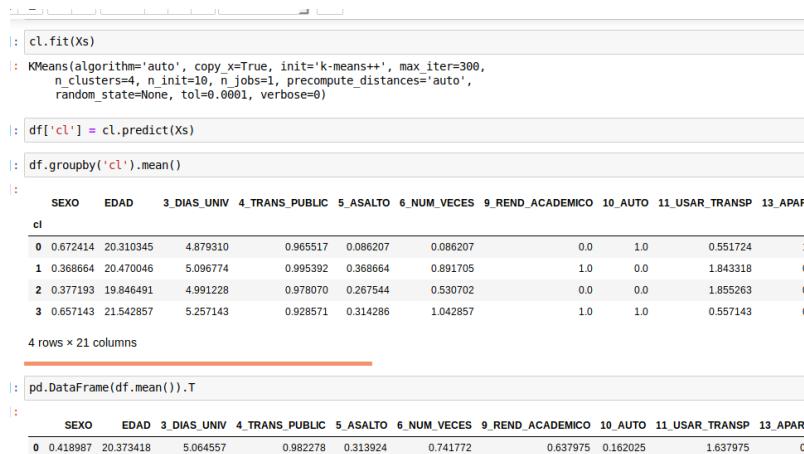
**Figura 7.7:** Limpieza Datos Parte - 7

Comenzamos el tratamiento para realizar los clusters, en total hicimos 4, para ver el perfil de nuestros encuestados, nos dimos cuenta que tenemos estudiantes que nunca han sido asaltados, que cuentan con auto propio y que no estarían dispuestos a utilizar el transporte público.



**Figura 7.8:** Limpieza Datos Parte - 8

Por otro lado, también tenemos población que no cuenta con auto propio, que afecta en su rendimiento escolar, que han sido asaltados demasiadas veces y por ende estarían dispuestos a utilizar nuestro servicio.



**Figura 7.9:** Limpieza Datos Parte - 9

Por último, con el tratamiento ya dado a los datos, obtamos por realizar una regresión y predecir si el alumnado estaría dispuesto a utilizar nuestro servicio o no, considerando ciertas variables explicativas. El código de la regresión se muestra en el siguiente apartado. Cabe destacar, que el modelo se ajustó muy bien a nuestros datos, es decir, la implementación de Machine Learning fue la adecuada.

# CAPÍTULO 8

## CÓDIGO-REGRESIÓN

The screenshot shows a Jupyter Notebook interface with several code cells. The first cell (In [1]) imports various Python libraries including pandas, numpy, and machine learning models from sklearn. The second cell (In [2]) starts with a comment '#Aquí cargo la base'. The third cell (In [5]) reads a CSV file named 'BaseTemas.csv' using pd.read\_csv(). The fourth cell (In [6]) contains two comments: '#Aquí le quito las personas que dicen que no quieren utilizar nuestro servicio' and '#Aquí no recuerdo qué número representa los que no'. The fifth cell (In [7]) filters the data to exclude rows where '11\_USAR\_TRANSPI' is not equal to 0. The sixth cell (In [8]) displays the head of the cleaned dataset.

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.model_selection import train_test_split
import seaborn
%matplotlib inline

#Aquí cargo la base

encuesta_df= pd.read_csv('BaseTemas.csv', encoding='utf8')

#Aquí le quito las personas que dicen que no quieren utilizar nuestro servicio
#Aquí no recuerdo qué número representa los que no

base_limpia=encuesta_df[encuesta_df['11_USAR_TRANSPI']!=0]

base_limpia.head()
```

Figura 8.1: Regresión parte 1

The screenshot shows a Jupyter Notebook interface with code cells and a resulting table. The first cell (In [7]) filters the data to exclude rows where '11\_USAR\_TRANSPI' is not equal to 0. The second cell (In [8]) displays the head of the cleaned dataset. The resulting table (Out[8]) shows 5 rows of data across 38 columns. The columns include SEXO, EDAD, CARRERA, DM, 1\_TRANSPI, 2\_TIEM\_DEMORA, 3\_DIAS\_UNIV, 4\_TRANS\_PUBLIC, 5\_ASALTO, 6\_NUM\_VECES, and others. The data includes details like travel mode (Microbus, Bicicleta, Metro, Automovil propio, Microbus, A pie, caminando), age (21, 20, 18, 23), and gender (F).

	SEXO	EDAD	CARRERA	DM	1_TRANSPI	2_TIEM_DEMORA	3_DIAS_UNIV	4_TRANS_PUBLIC	5_ASALTO	6_NUM_VECES	...	25
0	1	21	Actuaría	Acambay	Microbus, Bicicleta, Metro	04:00hrs	5	1	0	0	...	F
1	0	20	Actuaría	Acolmán	Microbus	01:00hrs	6	1	0	0	...	F
2	1	18	Actuaría	Acolmán	Automóvil propio, Microbus, A pie (caminando)...	01:00hrs	5	1	1	1	...	F
3	1	23	Actuaría	Álvaro Obregón	Microbus, Metro, Combi	02:30hrs	6	1	0	0	...	F
4	1	20	Actuaría	Álvaro Obregón	Microbus, A pie (caminando), Metro	02:30hrs	6	1	0	0	...	

5 rows × 38 columns

Figura 8.2: Regresión parte 2

	SEXO	EDAD	CARRERA	DM	1_TRANSPI	2_TIEM_DEMORA	3_DIAS_UNIV	4_TRANS_PUBLIC	5_ASALTO	6_NUM_VECES	...	25
0	1	21	Actuaría	Acambay	Microbús, Bicicleta, Metro	04:00hrs	5	1	0	0	...	1
1	0	20	Actuaría	Acolman	Microbús	01:00hrs	6	1	0	0	...	1
2	1	18	Actuaría	Acolman	Automóvil propio, Microbús, A pie (caminando)...	01:00hrs	5	1	1	1	...	1
3	1	23	Actuaría	Álvaro Obregón	Microbús, Metro, Combi	02:30hrs	6	1	0	0	...	1
4	1	20	Actuaría	Álvaro Obregón	Microbús, A pie (caminando), Metro	02:30hrs	6	1	0	0	...	1

5 rows × 38 columns

Figura 8.3: Regresión parte 3

	SEXO	EDAD	CARRERA	1_TRANSPI	2_TIEM_DEMORA	3_DIAS_UNIV	4_TRANS_PUBLIC	5_ASALTO	6_NUM_VECES	7_S
DM										
Acambay	1	1	1	1	1	1	1	1	1	1
Acolman	2	2	2	2	2	2	2	2	2	2
Almoloya de Juárez	1	1	1	1	1	1	1	1	1	1
Atizapán	2	2	2	2	2	2	2	2	2	2
Atizapán de Zaragoza	42	42	42	42	42	42	42	42	42	42
Azcapotzalco	17	17	17	17	17	17	17	17	17	17
Benito Juárez	9	9	9	9	9	9	9	9	9	9
Chalco	2	2	2	2	2	2	2	2	2	2
Chapultepec	1	1	1	1	1	1	1	1	1	1
Chimalhuacán	4	4	4	4	4	4	4	4	4	4
Coacalco de Berriozábal	15	15	15	15	15	15	15	15	15	15
Coyoacán	9	9	9	9	9	9	9	9	9	9
Coyotepec	2	2	2	2	2	2	2	2	2	2
Cuajimalpa de										

Figura 8.4: Regresión parte 4

	SEXO	EDAD	CARRERA	1_TRANSPI	2_TIEM_DEMORA	3_DIAS_UNIV	4_TRANS_PUBLIC	5_ASALTO	6_NUM_VECES	7_PRO
DM										
Acambay	1	1	1	1	1	1	1	1	1	1
Acolman	2	2	2	2	2	2	2	2	2	2
Almoloya de Juárez	1	1	1	1	1	1	1	1	1	1
Atizapán	2	2	2	2	2	2	2	2	2	2
Atizapán de Zaragoza	42	42	42	42	42	42	42	42	42	42
Azcapotzalco	17	17	17	17	17	17	17	17	17	17
Benito Juárez	9	9	9	9	9	9	9	9	9	9
Chalco	2	2	2	2	2	2	2	2	2	2
Chapultepec	1	1	1	1	1	1	1	1	1	1
Chimalhuacán	4	4	4	4	4	4	4	4	4	4
Coacalco de Berriozábal	15	15	15	15	15	15	15	15	15	15
Coyoacán	9	9	9	9	9	9	9	9	9	9
Coyotepec	2	2	2	2	2	2	2	2	2	2
Cuajimalpa de										

Figura 8.5: Regresión parte 5

```

In [31]: base_limpia.y.value_counts(normalize=True)
Out[31]: True    0.937861
          False   0.062139
          Name: y, dtype: float64

In [32]: y = base_limpia['y'].copy()

In [33]: lr = LogisticRegression()

In [34]: Xt,Xv,yt,yv =train_test_split(X,y,train_size=0.7)
         lr.fit(Xt,yt)

         print "train %.3f"%accuracy_score(y_pred=lr.predict(Xt),y_true=yt)
         print "validate %.3f"%accuracy_score(y_pred=lr.predict(Xv),y_true=yv)

         train 0.940
         validate 0.933
/usr/lib/python2.7/dist-packages/sklearn/model_selection/_split.py:2026: FutureWarning: From version 0.21,
l always complement train_size unless both are specified.
FutureWarning

In [36]: lr.predict(np.array([2,6,10]).reshape(1, -1))[0]
Out[36]: True

```

**Figura 8.6:** Regresión parte 6

```

In [36]: lr.predict(np.array([2,6,10]).reshape(1, -1))[0]
Out[36]: True

In [37]: lr.coef_
Out[37]: array([0.45789474, 0.35750166, 0.04658682])

In [38]: lr.intercept_
Out[38]: array([0.21212462])

In [39]: import pickle

In [40]: pickle.dump(lr,open('modelo.modelo','wb'))

In [41]: np.array([0])[0]
Out[41]: 0

In [42]: from sklearn.feature_selection import SelectKBest
In [46]: sk = SelectKBest(k=1)
In [47]: sk.fit(X,y)
Out[47]: SelectKBest(k=1, score_func=<function f_classif at 0x7fafcdd2da28>)

```

**Figura 8.7:** Regresión parte 7

Con el código ya mostrado, notamos que el modelo se ajusta muy bien a nuestros datos.



---

# CAPÍTULO 9

---

## DICCIONARIO DE DATOS

**Cuadro 9.1:** Diccionario de Datos

No.	NOMBRE DEL CAMPO	TIPO	DESCRIPCION DEL CAMPO	RANGO	LONGITUD	OBSERVACIONES
1	SEXO	BINARIA	Sexo	0-1	1	0-Hombre/1-Mujer
2	EDAD	NUMÉRICO	Edad	15-49	2	-
3	CARRERA	ALFABÉTICO	Carrera	-	-	-
4	DM	ALFABÉTICO	Delegación/Municipio	-	-	-
5	1_TRANS	ALFABÉTICO	1.-¿Cuál es el medio de transporte que usa frecuentemente para trasladarse a la FES Acatlán? (Puede seleccionar más de una).	-	-	-
6	2_TDEMORA	ALFANUMÉRICO	2.-¿Cuánto tiempo (minutos) aproximadamente demora en llegar desde su lugar de residencia a la FES Acatlán?	00:20 hrs - 05:00 hrs	-	-
7	3_DIAS_UNIV	NUMÉRICO	3.-¿Cuántas veces a la semana viene a la Universidad?	0-7	1	-
8	4_TRANS_PUBLIC	BINARIA	4.-¿Ha utilizado alguna vez el transporte público?	0-1	1	0-No/1-Sí
9	5_ASALTO	BINARIA	5.-¿Has sufrido algún asalto rumbo a la Universidad?	0-1	1	0-No/1-Sí
10	6_NUM_VECES	NUMÉRICO	6.-En caso de que tu respuesta haya sido "si" contesta la sig. pregunta. ¿Cuántas veces?	0-25	2	-
11	7_PROBL_TRANSPI	ALFABÉTICO	7.-De los siguientes problemas, selecciona aquellos que padeciste en el transporte público rumbo a la Universidad.	-	-	-
12	8_LLEG_TARDE_UNIV	ALFABÉTICO	8.-¿Has llegado tarde a alguna de tus clases debido a algún problema con el transporte público?	-	-	-

No.	NOMBRE DEL CAMPO	TIPO	DESCRIPCION DEL CAMPO	RANGO	LONGITUD	OBSERVACIONES
13	9_REND_ACADEMICO	BINARIA	9.-¿Piensas que el transportante de la escuela a tu casa es un problema que afecta a tu calidad de estudiante?	0-1	1	0-No/1-Sí
14	10_AUTO	BINARIA	10.-¿Cuentas con auto? (Que solo sea de uso exclusivo para ti)	0-1	1	0-No/1-Sí
15	11_USAR_TRANSPI	BINARIA	11.-En caso de que tu respuesta haya sido "sí", ¿Contemplarás el usar transporte público? (Si tu respuesta fue "No" pasa a la pregunta 12)	0-2	1	0-No/1-Sí/2-No sabe
16	12_RAZONES	ALFABÉTICO	12.-En caso de que tu respuesta haya sido "sí", ¿Por cuál de las siguientes razones lo usarías? (Puedes elegir más de una)	-	-	-
17	13_APAR_ELECT	BINARIA	13.-¿En tu día a día en la escuela, llevas contigo algún aparato electrónico como computadoras, tablets, celulares, etc?	0-1	1	0-No/1-Sí
18	14_RAZON_TRANSPI	ALFABÉTICO	14.-¿Cuál es la principal razón por la que eliges tu medio de transporte? (Puedes elegir más de una).	-	-	-
19	15_FREC_TRANSPI	ALFABÉTICO	15.-¿Qué tan seguido utilizas el transporte público?	-	-	-
20	16_CALIF_TRANSPI	ALFABÉTICO	16.-¿Cómo califica la limpieza y condiciones del transporte público que usas?	-	-	-
21	17_CALIF_CONDUC	ALFABÉTICO	17.-¿Cómo calificas el trato de los conductores del transporte público?	-	-	-
22	18_MEJORAS	ALFABÉTICO	18.-¿En qué crees que debería mejorar el transporte público? (Puedes elegir más de una)	-	-	-
23	19_REDUC_TRAF	BINARIA	19.-¿Crees que el transporte público ayuda a tener menos tráfico?	0-1	1	0-No/1-Sí
24	20_ENTRADA	ALFANUMÉRCICO	20.-Normalmente ¿A qué hora entras a la Universidad?	-	-	-
25	21_SALIDA	ALFANUMÉRCICO	21.-Normalmente ¿A qué hora sales de la Universidad?	-	-	-
26	22_MIN_ESPERA	NUMÉRICO	22.-En promedio ¿Cuántos minutos esperas tu transporte para ir a la Universidad?	0-90	2	-
27	23_HORA_INSEG	ALFANUMÉRCICO	23.-¿A qué hora crees que hay más inseguridad en el transporte público en el que viajas?	-	-	-
28	24_FALTAS_CLASE	BINARIA	24.-Has faltado a clases porque el traslado no vale la pena ?	0-1	1	0-No/1-Sí
29	25_HUMOR	ALFABÉTICO	25.-¿Cuál es tu humor después de terminar tu traslado?	-	-	-
30	26_TRANS_PRIV	BINARIA	26.-¿Te gustaría un transporte privado propio de la Universidad?	0-1	1	0-No/1-Sí

No.	NOMBRE DEL CAMPO	TIPO	DESCRIPCION DEL CAMPO	RANGO	LONGITUD	OBSERVACIONES
31	27_APPLICACION	BINARIA	27.-¿Conoces alguna aplicación móvil que te permita hacer uso de algún medio de transporte?	0-1	1	0-No/1-Sí
32	28_NOMB_APPLIC	ALFABÉTICO	28.-En caso de que tu respuesta haya sido "Sí" contesta la sig. pregunta, ¿Cuál?	-	-	-
33	29_TRANS_UTIL	BINARIA	29.-¿Crees que un transporte privado para estudiantes con rutas estratégicas sería útil para los estudiantes de la FES Acatlán?	0-1	1	0-No/1-Sí
34	30_CREDENCIAL	BINARIO	30.- ¿Consideras que el solicitar credencial/ tira de materias para subir a un transporte privado sea suficiente para asegurarnos de que dicho transporte sea seguro?	0-1	1	0-No/1-Sí
35	31_FACT_TRANS	ALFABÉTICO	31.-¿Cuáles de los siguientes factores te convencen para ser usuario de un transporte público? (Puedes elegir más de una)	-	-	-
36	32_SEGURIDAD	BINARIA	32.-¿Te sentirías más seguro si el transporte que brinda la Universidad al metro fuese sólo para estudiantes?	0-1	1	0-No/1-Sí
37	LATITUD	NUMÉRICO	Latitud			
38	LONGITUD	NUMÉRICO	Longitud			



---

# CAPÍTULO 10

---

## CONCLUSIONES

El desarrollo de las tecnologías ha rebasado todas nuestras expectativas. Propuestas de modelos tecnológicos que veíamos hace un par de años solo en películas de ciencia y ficción como algo poco probable de suceder, hoy en día son herramientas indispensables en el desenvolvimiento laboral y social de las personas.

A pesar de todos estos grandes avances tecnológicos de los cuales hemos sido testigos durante los últimos años, en la actualidad, aún no se cuenta con alguna herramienta que pueda darnos a conocer con certeza que pasará en el futuro. Es aquí, donde el desarrollo de dichas tecnologías, busca subsanar ese vacío que deja esa incertidumbre. ¿Cómo?, a través de los datos.

Los datos son una herramienta muy poderosa, ya que mediante ellos, podemos conocer y tratar predecir el comportamiento de ciertos fenómenos que son de nuestro interés, ya sea como estrategia de venta, inversión, planeación, etc., a nivel individual, o mejor aun, dentro de una corporación.

Sin duda alguna, la correcta explotación de los datos son la mejor representación de la materia prima de las industrias del futuro. Y no solo del futuro, ya que la mayoría de las grandes empresas en la actualidad, han alcanzado todo ese éxito gracias a los datos.

Así pues, en la actualidad, indudablemente, en el desarrollo laboral completo de un profesional, el dominio de dichas tecnologías tiene que ser parte esencial. Y no solo a nivel individual, sino también hablando de una organización completa. Una empresa que ya que conforme avanzamos en el tiempo, la competencia y la exigencia se hace mas notable en el ámbito laboral.



---

# BIBLIOGRAFÍA

- [1] **Josep Lluís Cano.**, (2007). *BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN*. ESADE Business School, 1, 393. 2018, De ESADE Base de datos.
- [2] JACK FLEITMAN. (2010). *LA IMPORTANCIA DE LOS TABLEROS DE CONTROL*. «Libro evaluación integral para implantar modelos de calidad», De Ciemsa Base de datos.
- [3] **Josep Lluís Cano**, (2007). *BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN*. ESADE Business School, 1, 393. 2018, De ESADE Base de datos.
- [4] J. AGUILAR, I. BESSEMBEL, M. CERDA, F. HIDROBO, F. NARCISO. (2008). ,*Una Metodología para el Modelado de Sistemas de Ingeniería Orientado a Agentes*. 13 de Noviembre del 2018, de Dialnet Sitio web: <https://dialnet.unirioja.es/servlet/articulo?codigo=2856317>.
- [5] FERNANDO SANCHO CAPARRINI.(2017)., «Clasificación Supervisada y No Supervisada.», 13 de Noviembre del 2018, , Sitio web: <http://www.cs.us.es/> fsancho/?e=77fbclid=IwAR00OEZMjeAOLbGYvf-DU6b72EU0AOH3g4Fu7hjhxXhSFzemjxOA8NrT5Ms
- [6] *Tech Target*. (2017). Guía de Gestión de Proyectos de TI: Desde Scrum hasta la Medición del ROI. 2017, de Serch Data CenterSitiweb:<https://searchdatacenter.techtarget.com/es/definicion/Base-de-datos-relacional-Campus-MVP>. (2014).
- [7] **Josep Cano**, (2008). *BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN*. ESADE Business School, 1, 393. 2018, De ESADE Base de datos.
- [8] Breiman, L. (junio de 1997). „Arcing The Edge” (PDF) . Informe técnico 486 . Departamento de Estadística, Universidad de California, Berkeley.  
John Goddard. (2000). un algoritmo para el entrenamiento de maquinas de vector soporte para regresión. Revista de Matematica: Teoría y Aplicaciones, 7, 107-108.