

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE

DEPARTAMENTO DE ELECTRÓNICA, SISTEMAS E INFORMÁTICA
ESPECIALIDAD EN SISTEMAS EMBEBIDOS

MSC2526A – PROGRAMACIÓN PARA ANÁLISIS DE DATOS
PRIMAVERA 2021



ITESO, Universidad
Jesuita de Guadalajara

ANÁLISIS DE CARTERA DE CLIENTES

PROYECTO FINAL

Presenta:

*Dorian Ricardo Martinez Preciado,
744445, dorianr.martinez@iteso.mx*

GitHub del proyecto

<https://github.com/DorianMtzP/ProyectoAnalisisDatos>

Tlaquepaque, Jalisco. Mayo de 2023.

1.RESUMEN

Se presenta una breve introducción al análisis de la cartera de clientes de una empresa dedicada a los generadores eléctricos ubicada en California. El análisis tiene como objetivo principal ayudar en la toma de decisiones para el futuro de la empresa, la cual está buscando cambiarse de su ubicación actual con la meta final de lograr expandirse, en el proceso de expansión, se requerirá contratar y entrenar nuevos técnicos con diferentes tipos de habilidades.

El conocer la ubicación de los clientes nos ayudara a concentrar nuestra búsqueda del lugar para la nueva oficina, ya que esta debe de ser ubicada estratégicamente cerca de donde se encuentre la mayoría. En cuanto a las relaciones entre la cantidad de clientes, el nivel técnico que requieren, su prioridad asignada al igual que su ubicación, nos ayudara a resolver otras variables como, cuales son las mejores fechas para contratar y entrenar a los técnicos, en que áreas nos conviene que residan dichos técnicos, es decir, si residen cerca de la nueva oficina o cerca de donde hay una mayor cantidad de clientes, lo mas probable es que nuestro tiempo de respuesta a llamadas de emergencia mejore.

También encontramos que existe una relación entre el tipo de técnico y la fecha de los mantenimientos, lo cual nos puede ayudar con las prioridades al tiempo de contratación dependiendo de la fecha en que se lanza la campaña.

2.TABLA DE CONTENIDOS

1. RESUMEN.....	2
3. OBJETIVO DE INVESTIGACIÓN.....	4
3.1. INTRODUCCIÓN.....	5
3.2. ANTECEDENTES	5
3.3. JUSTIFICACIÓN.....	5
3.4. PROBLEMA	5
3.5. OBJETIVOS.....	5
3.6. NOVEDAD CIENTÍFICA, TECNOLÓGICA O APORTACIÓN	5
4. OBTENCIÓN DE LOS DATOS.....	6
4.1. PROCESO DE OBTENCIÓN DE LOS DATOS	7
4.2. CONCLUSIONES	8
5. PREPARACIÓN DE LOS DATOS	9
5.1. PROCESO DE PREPARACIÓN DE LOS DATOS	10
5.2. CONCLUSIONES	11
6. PRESENTACIÓN DE RESULTADOS	12
6.1. PRESENTACIÓN DE RESULTADOS	13
7. CONCLUSIONES.....	18
7.1. CONCLUSIONES	19
7.2. TRABAJO A FUTURO	20
8. REFERENCIAS BIBLIOGRAFICAS	21

3.OBJETIVO DE INVESTIGACIÓN

Resumen: *En esta sección se presenta el objetivo de la investigación elaborada. Habla sobre la empresa de donde provinieron los datos, y su necesidad de cambiar de domicilio y expandirse realizando una mayor contratación de personal. Los datos analizados ayudarán a encontrar la relación entre estas variables y asistirán con la decisión.*

3.1. Introducción

Este documento pretende dejar evidencia del análisis realizado a la base de datos de la empresa de generadores eléctricos.

3.2. Antecedentes

La empresa se encuentra ubicada en la ciudad de Hayward, CA. En una zona céntrica de la ciudad, debido a la actual ubicación le es imposible expandirse, por lo cual se busca mudar las oficinas a una nueva localidad. Esto implica una serie de cambios en los cuales contar con información puntual ayudara a tomar la mejor decisión.

3.3. Justificación

El expandirse implica contratar una mayor cantidad de personal técnico, el cual es catalogado por niveles de conocimiento y habilidades, yendo del 1 al 3 y MC. Es aquí donde esta investigación entra en juego ya que nos ayudara a saber de que ciudades nos conviene contratar el personal (esto es importante para las llamadas de emergencia) así como también saber qué porcentaje de técnicos de cada nivel debemos tener para poder brindar servicio de manera pronta y eficaz a los clientes. El saber cuáles son las fechas con mayor actividad en la empresa nos ayudara a fijar un calendario de entrenamientos y vacaciones.

3.4. Problema

Analizar la información existente para determinar el plan de acción óptimo para el crecimiento y migración de la empresa a su nueva locación.

3.5. Objetivos

3.5.1. Objetivos Específicos:

1. Determinar la cantidad de clientes por ciudad y encontrar las ciudades con mayor cantidad.
2. Encontrar el tipo predominante de clientes (Prioridad).
3. Carga laboral por mes.
4. Cantidad de clientes por tipo de técnico.
5. Relación que existe entre ubicación del cliente y su tipo (Prioridad).
6. Relación que existe entre ubicación del cliente y el tipo de técnico que necesita.

3.6. Novedad científica, tecnológica o aportación

La plataforma que la empresa utiliza para manejar sus clientes no cuenta con reportes que ayuden a analizar los datos, los reportes proporcionados son simplemente fragmentos exportados de la base de datos los cuales deben de ser posteriormente trabajados para su análisis. Este proyecto aportara al análisis de dichos datos, ya que en un futuro se podrá actualizar la base de datos y ejecutar el análisis nuevamente.

4.OBTENCIÓN DE LOS DATOS

Resumen: *En esta sección se presenta la metodología para la obtención de los datos utilizados. La base de datos a utilizar en este proyecto proviene de un pequeño reporte real de la lista de clientes de una empresa dedicada al mantenimiento, instalación, venta y reparación de generadores eléctricos (plantas de luz). El reporte fue creado con la intención de ser lo suficiente informativo como para poder trabajar con él sin tener que exponer datos críticos que pudieran afectar la privacidad de la empresa.*

4.1. Proceso de obtención de los datos

El CRM de la empresa cuenta con la funcionalidad de crear reportes, los reportes se dividen en varias categorías como lo son “trabajos”, “clientes”, “facturación”, entre otros. Dependiendo del tipo de reporte se muestran ciertos campos de información, es decir, un reporte de trabajos no mostrara la dirección de facturación del cliente, mas si proporcionara la dirección de servicio.

Para la base de datos se selecciono un tipo de reporte de “Clientes” el cual proporciona la suficiente información para obtener los resultados que deseamos. Cabe destacar que se omitió una gran cantidad de datos, esto con el fin de no violar la privacidad de la empresa. Se omitieron nombres, datos fiscales, direcciones (excepto por la ciudad), datos de los equipos, costos, etc.

Una vez configurado el reporte se exporto en un archivo CSV con longitud de 1818 filas y 8 columnas de datos, las cuales se describen a continuación:

- **CustomerNumber:**
 - Número único de identificación del cliente.
- **CustomerServiceZipCode:**
 - Código postal de la dirección donde se encuentra el generador eléctrico.
- **CustomerServiceCity:**
 - Ciudad donde se encuentra el generador eléctrico.
- **Priority:**
 - Prioridad que tiene el cliente, va del 1 al 8, siendo el 1. Life Safety el de mayor prioridad.
Los valores disponibles son:
 1. Life Safety
 2. Gold
 3. Silver
 4. Bronze
 5. Normal
 6. Residencial
 7. Remote
 8. Portable
- **TechType:**
 - Contiene el tipo de técnico requerido para realizar el mantenimiento.
Los valores disponibles son:
 1. Level 1
 2. Level 2
 3. Level 3
 4. MC
- **NextOilChangeDate:**
 - Fecha del próximo mantenimiento.
- **LastOilChangeDate:**
 - Fecha del último mantenimiento realizado.

- **CustomerIsActive:**
 - Indica si la cuenta del cliente se encuentra activa o no, si esta inactiva significa que el cliente ya no utiliza nuestros servicios.

4.2. Conclusiones

Se logro obtener la base de datos aprobada por la empresa para el trabajo a realizar, la cual contiene suficiente información para obtener los resultados deseados, sin exponer datos críticos.

5. PREPARACIÓN DE LOS DATOS

Resumen: En esta sección se presenta la manera en que la base de datos fue acondicionada para poder realizar el análisis. Se realizó una eliminación de espacios vacíos, al igual que de filas que no contenían toda la información requerida. También se descartaron los clientes que estaban inactivos, es decir, clientes que ya no utilizaban nuestros servicios. Se les dio formato a las columnas cambiando tipos de datos y regularizando la información contenida en ellas.

5.1. Proceso de preparación de los datos

Se encontraron los siguientes problemas base de datos:

- Existen varios clientes que estaban inactivos (ya no utilizan nuestros servicios)
- Hay valores nulos en la columna de “CustomerServiceZipCode” así como también hay algunos códigos postales clásicos y otros en versión extendida.
- El campo “TechType” no es consistente, tenía dos tipos de cadena para indicar el mismo valor
- Algunos clientes tienen valores nulos en sus columnas “CustomerServiceCity” y “Priority”
- La Prioridad y el Nivel técnicos son cadenas de texto, para realizar la matriz de correlación será necesario convertirlos a enteros
- Solo nos interesa el mes de la fecha del cambio de aceite más reciente.

A continuación, se muestran las columnas que requirieron algún tipo de tratamiento de datos. La preparación de datos se realizó en un Jupiter Notebook el cual puede ser encontrado en <https://github.com/DorianMtzP/ProyectoAnalisisDatos> [1]

CustomerNumber

El primer paso de la preparación fue establecer el índice de los datos, para esto utilice el único número que es único en el data frame, el número de cliente.

CustomerServiceZipCode

Se realizó la normalización de los códigos postales, tomando únicamente los primeros 5 dígitos, con esto se logró convertirlos todos en modo clásico, posteriormente se le hizo un cast a la columna para que esta fuera de tipo entero. También se eliminaron las filas con valores nulos.

Priority

Al inicio del análisis se utilizó sin ninguna otra modificación que la eliminación de las filas con valores nulos, posteriormente, para la creación del gráfico de correlación, se creó una copia y se extrajo solo la parte numérica del valor, para con esto obtener una escala numérica del 1 al 8.

TechType

El menú dropdown de la aplicación que llena la columna de "TechType" fue actualizado en algún punto en la vida de la empresa después de que ya había clientes en la base de datos, es por ello que algunas filas muestran solo el nivel del técnico (Level 2) y otras muestran el nivel y un ejemplo en paréntesis (Level 2 (i.e. liquid cooled)), para nuestros fines prácticos nos conviene ecualizar la información y eliminar los campos vacíos.

Para la primera parte del análisis simplemente se eliminaron las columnas con valores nulos y se mapearon los datos a una cadena de tipo “Level x” siendo “x” los niveles del 1 al 3. Para la segunda parte del análisis, se tomó el valor numérico y se casteó el valor a tipo entero.

NextOilChangeDate y LastOilChangeDate

Se realizó una conversión de los datos a tipo date time. Posteriormente se rellenaron los espacios vacíos realizando un remplazo utilizando los valores de ambas columnas y manteniendo la fecha más actual. De esta forma, en las filas donde existía al menos una fecha, se eliminaron los campos nulos y en las filas restantes donde aún existían fueron eliminadas.

Al realizar el proceso antes mencionado ambas columnas terminaron teniendo los mismos datos, por lo cual se procedió a eliminar una y a renombrar la otra, el nombre de la nueva columna es: “MaintenanceDate”.

En la segunda parte del análisis, donde se realiza la gráfica de correlación, se extrajo el mes, se convirtió a tipo entero y se guardó en una nueva columna “Month”

CustomerIsActive

Se filtraron las filas donde el valor era “False” y se eliminaron junto con los valores nulos.

5.2. Conclusiones

Esta parte del acondicionamiento de la base de datos fue la más retardadora de todo el proyecto. Después de haber invertido una considerable cantidad de tiempo, se logró tener una data frame que nos serviría para procesar los análisis requeridos.

6. PRESENTACIÓN DE RESULTADOS

Resumen: *En esta sección se presentan los resultados obtenidos del análisis de los datos. Se crearon gráficos que ayudaron a comprender la relación entre los datos analizados y ayudaran a la toma de decisiones en un futuro cercano.*

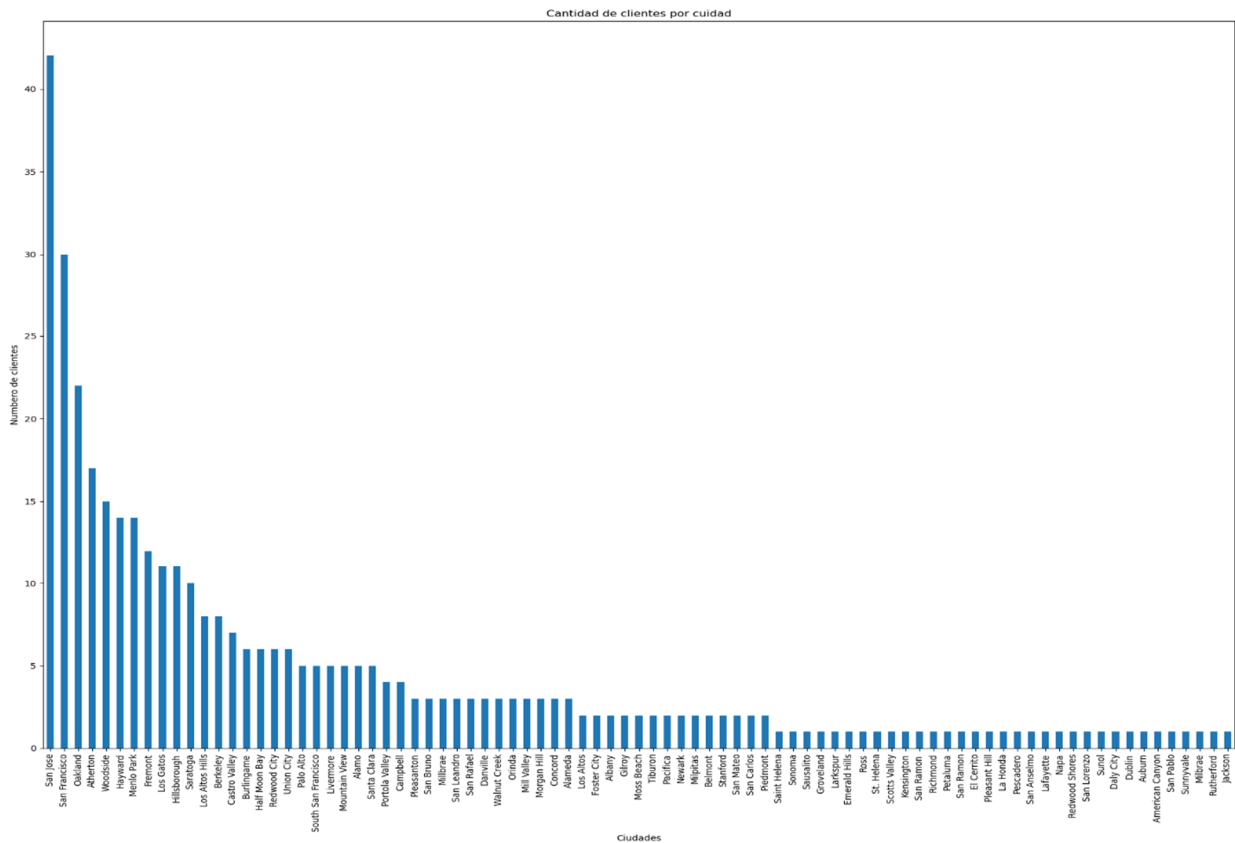
6.1. Presentación de resultados

Mediante la creación de graficas se logro analizar los datos trabajados a continuación se muestran los resultados listando los objetivos específicos iniciales y su resultado obtenido

Determinar la cantidad de clientes por ciudad y encontrar las ciudades con mayor cantidad.

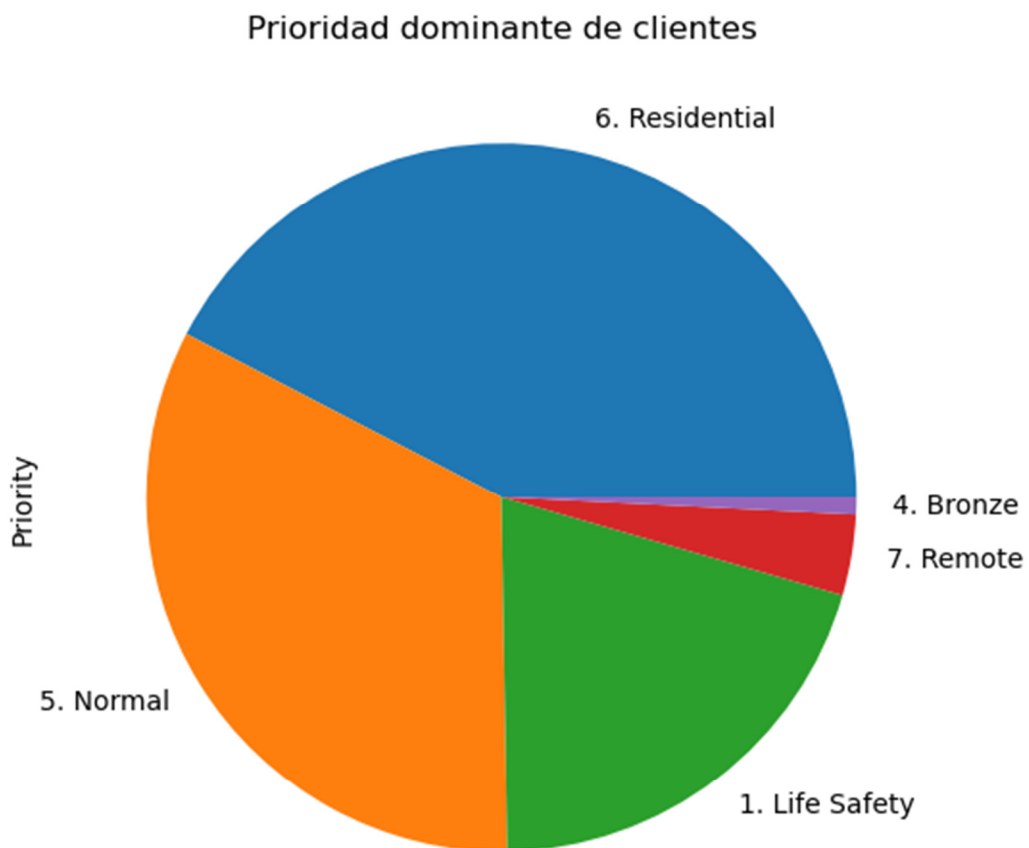
Se determino que las 5 ciudades con mayor cantidad de clientes son:

1. San José
2. San Francisco
3. Oakland
4. Atherton
5. Woodside



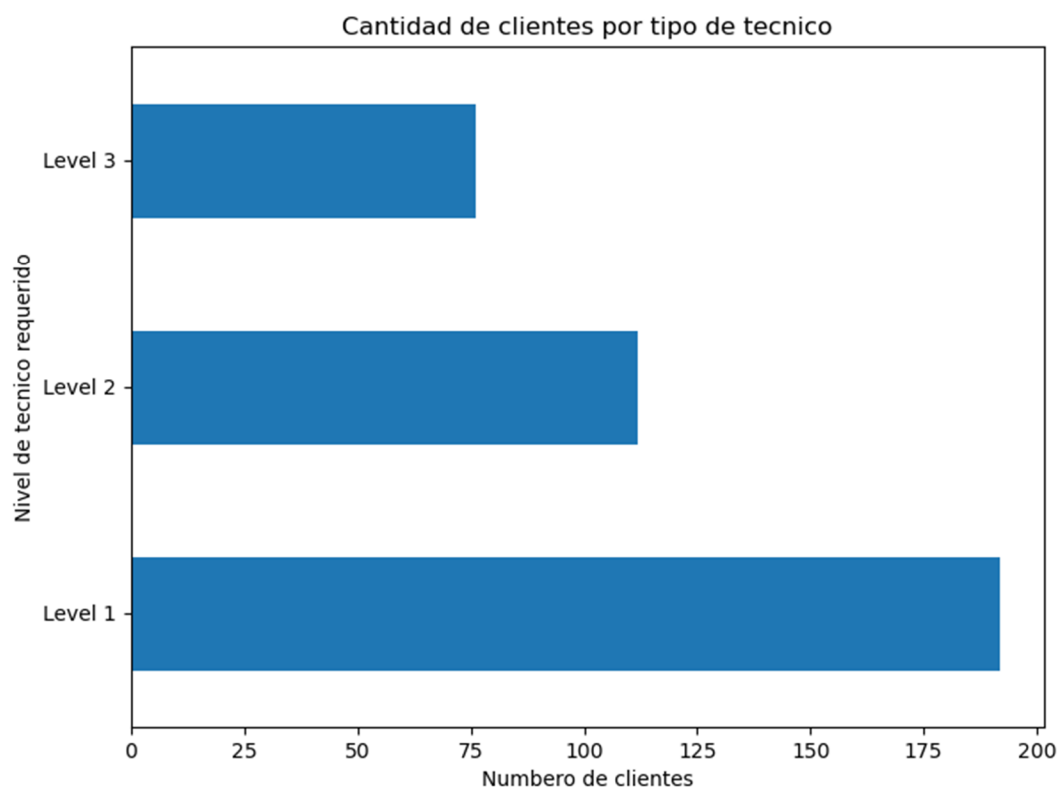
Encontrar el tipo predominante de clientes (Prioridad)

Se puede observar que el data frame no contiene datos de algunas de las categorías de prioridad. también se puede apreciar que poco menos de la mitad de nuestros clientes son de tipo residencial.



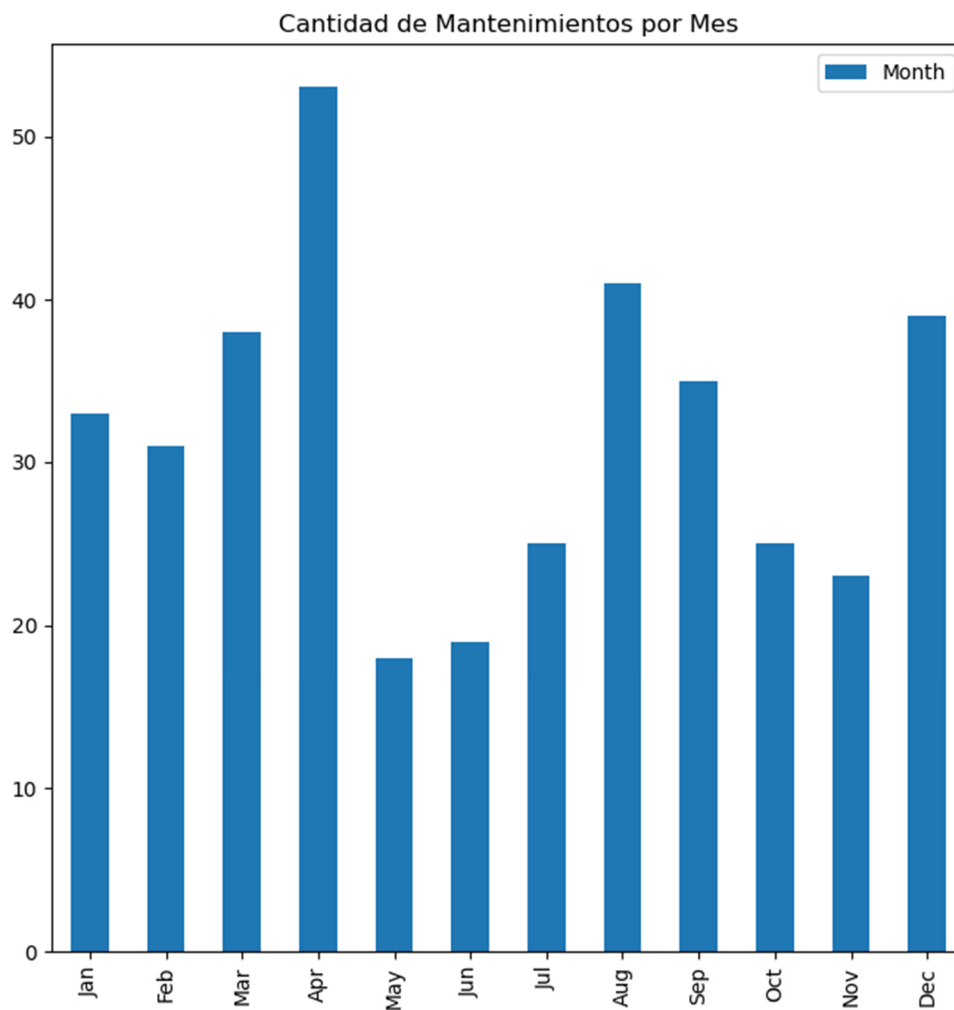
Cantidad de clientes por tipo de técnico.

Se observa que la mayor cantidad de clientes requieren un tipo de tecnico nivel 1, seguidos de nivel 2 y 3.



Carga laboral por mes

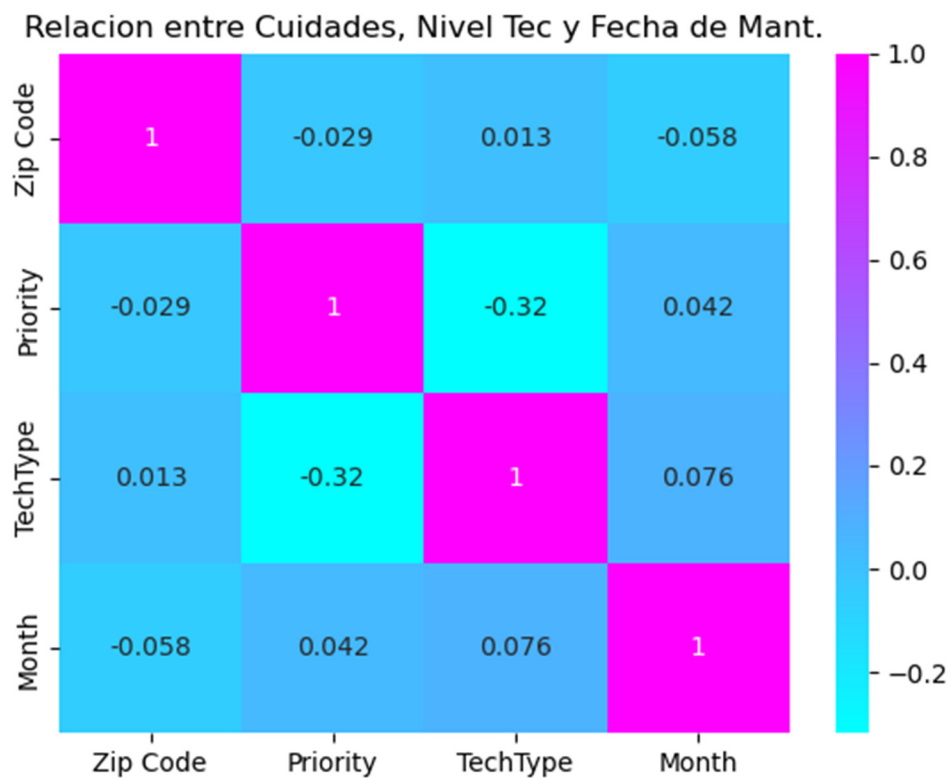
Por medio de una gráfica que muestra la distribución de mantenimientos a lo largo del año, se puede observar que los Meses con mas carga laboral son Abril, Agosto y Diciembre. Siendo el periodo que va de Mayo a Julio el de menor actividad.



Relación que existe entre ubicación del cliente y su tipo (Prioridad) así como también el tipo de técnico que necesita.

Para obtener estos datos se realizó un acondicionamiento de los datos adicional (segunda parte). Posteriormente se creó una gráfica de correlación, las dos correlaciones más fuertes son:

- 0.076 entre el mes de mantenimiento y el tipo de técnico
- 0.058 entre el mes de mantenimiento y la ubicación del equipo



7.CONCLUSIONES

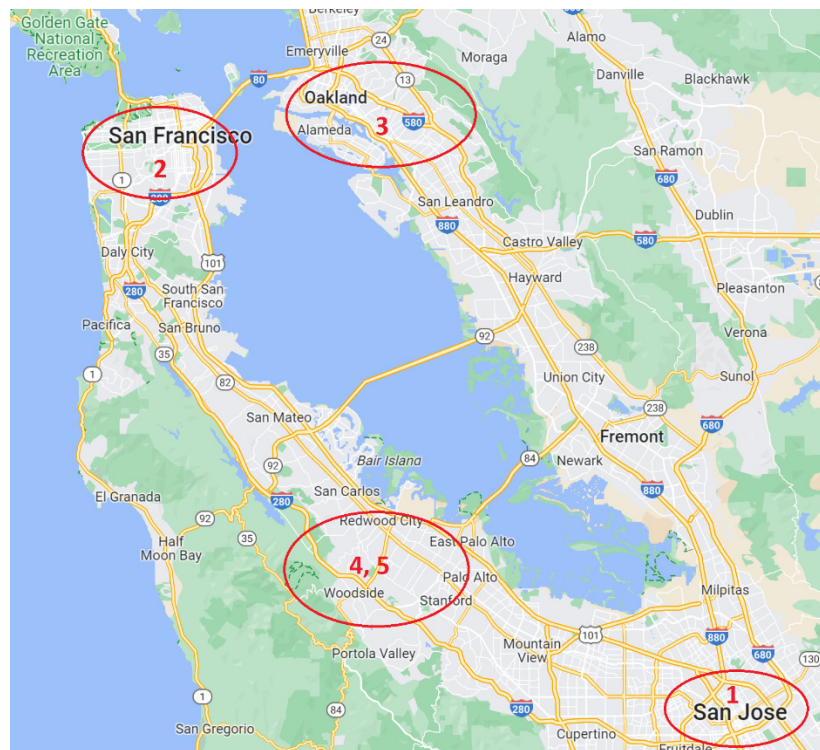
Resumen: *En esta sección se presentan las conclusiones obtenidas después de analizar los resultados obtenidos de cada objetivo específico planteados al inicio de este documento. También se describen los trabajos a futuro para mejorar el análisis realizado.*

7.1. Conclusiones

A continuación, se muestran las conclusiones a las que se llegó utilizando los resultados obtenidos, dichas conclusiones se listan por objetivo específico.

Objetivo 1. Determinar la cantidad de clientes por ciudad y encontrar las ciudades con mayor cantidad.

Al analizar los datos llegamos a la conclusión de que la ubicación de la nueva oficina, si estuviera cerca de *Redwood city*, seria la mejor para nuestra cartera de clientes actual. Mas sin embargo, debido a la forma geográfica de la área de la bahía, es posible que una oficina ubicada en *Union City* sea mas optima, ya que estaría cerda de las ciudades 1, 3,4 y 5 con mayor número de clientes y además daría acceso rápido a los clientes ubicados al este, los cuales están mas dispersos pero siguen siendo una cantidad considerable.



Objetivo 2. Encontrar el tipo predominante de clientes (Prioridad)

Los datos analizados en este objetivo pueden causar confusión, ya que si bien la gráfica muestra una gran cantidad de clientes residenciales, esto no significa que la mayoría de los equipos a los que damos servicio son de tipo residencial, ya que los clientes comerciales o industriales (5.Normal) suelen tener más de un generador por cuenta. Por lo cual podríamos concluir que este análisis es inconcluso, se necesitan mas datos para lograr un panorama real.

Objetivo 3. Cantidad de clientes por tipo de técnico.

Estos resultados también son un poco engañosos, debido a que cae en el mismo juego que la anterior. Los clientes de tipo residencial son asignados técnicos de "Nivel 1". En la práctica, los técnicos Nivel 2 son los más dinámicos ya que pueden atender a la mayoría de clientes comerciales e industriales. Por su parte, los clientes de "Live Safety" así como clientes especiales, requieren técnicos "Nivel 3", creo que este dato si es acertado.

Objetivo 4. Carga laboral por mes

Se pudo comprobar que la información reafirma lo que se ve empíricamente en la realidad. Excepto por los meses de Octubre y Noviembre los cuales se utilizan para darle mantenimiento a unos pocos clientes Industriales que tienen una enorme cantidad de generadores, pero que, en esta grafica cuentan solo como 1 elemento. Podemos concluir también que la mejor temporada para contratar técnicos y entrenarlos son los meses de Mayo a Julio. Mientras que la temporada con mas actividad es de Diciembre a Abril, lo cual concuerda con las fechas en que el frio y la lluvia suceden, dos factores que suelen recordar al cliente que necesita darle mantenimiento a su generador.

Objetivo 5 y 6. Relación que existe entre ubicación del cliente y su tipo (Prioridad) así como también el tipo de técnico que necesita.

Me sorprendió ver la alta relación entre el tipo de técnico y el mes de mantenimiento, después de pensarlo un poco me doy cuenta de que si es correcto, los meses de Octubre, Noviembre y Diciembre los dedicamos a cuentas Industriales y Live Safety, las cuales requieren técnicos de niveles 2 y 3, los meses de Abril y Marzo por lo regular es cuando los clientes Residenciales prefieren hacer sus mantenimientos. Me parece muy interesante esto de la correlación.

Originalmente esperaba que hubiera una mayor relación entre el tipo de técnico y el código postal, esto porque hay zonas residenciales (técnicos Nivel 1) y zonas industriales/comerciales (técnicos Nivel 2 y 3). La relación, si bien se muestra en segundo lugar, solo tiene un 0.058 comparado con el 0.076 de la correlación dominante del mes de mantenimiento vs técnico

7.2. Trabajo a Futuro

Dos de los objetivos no fueron alcanzados de manera satisfactoria debido a que la base de datos con contaba con dicha información. Convendría entonces, al menos para uso interno de la empresa, el mejorar la base de datos para lograr que un mayor porcentaje de clientes tuvieran sus datos completos, por lo menos, los requeridos para este análisis. El estado actual de la base de datos fue precario, de las 1818 líneas de datos se pudieron utilizar solo 380.

Otro elemento por considerar podría ser el realizar un reporte de los trabajos y unirlo al reporte de los clientes, esto requeriría un tratamiento de datos extenso, mas sin embargo se podría obtener información de otros tipos de servicios como lo son las reparaciones, reemplazos, garantías entre otros.

8. REFERENCIAS BIBLIOGRAFICAS

- [1] D. Martinez, «GitHub,» 11 05 2023. [En línea]. Available:
<https://github.com/DorianMtzP/ProjectoAnalisisDatos>. [Último acceso: 11 05 2023].
- [2] AAAAGen, «Index,» AAAA Generator Services, 2023. [En línea]. Available: aaaagen.com.
[Último acceso: 10 05 2023].