



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
INSTITUTO DE INGENIERÍA MATEMÁTICA Y COMPUTACIONAL
LIC. EN ING. EN CIENCIA DE DATOS
IMT-2200

La relación entre el transporte público en la región metropolitana y los diferentes índices de su población

Propuesta

Grupo 13

Integrantes:

Jeferson Alzate, Joaquín Bächler, Matias Morales, Franco Moya, Gabriel Vergel

Fecha de entrega: <día> de <mes> de <año>



Índice de contenido

1. Contexto y Motivación	2
2. Objetivos	2
3. Datos	3
4. Preguntas de Investigación	4
5. Diseño tentativo	5
5.1. Regresión Lineal	5



1. Contexto y Motivación

Dado que somos usuarios cotidianos del servicio de transporte público, es natural que tengamos un interés en él. A diario nos encontramos con vagones (o buses) llenos, vacíos o normales ¿Qué hay de los usuarios de esos vagones? ¿Hay algún patrón? ¿Alguna tendencia? ¿Siempre ha sido así?

Son preguntas que uno podría, ya sea por curiosidad, ya sea por planear mejor un viaje, ya sea por entender mejor, ya sea por lo que sea. La cuestión es que hay algo interesante que puede ser investigado.

Pensamos que de este proceso podremos, ojala, responder alguna de estas preguntas, y sino, quizá alguna nueva que ocurra durante el proceso de recopilación o trabajo con los datos.

Lo bueno que tiene esta iniciativa es que es útil para gran parte de la población, pues gran parte de ella hace uso del sistema de transporte público. Por otro lado, entendiendo que el órgano que regula este sistema debe estar informado acerca de su propio funcionamiento, podríamos decir que la información (más que la información, las conclusiones) podrían de servirle en algún grado.

2. Objetivos

A partir del proceso que nos proponemos llevar acabo podremos obtener relaciones entre :

- Rango etario y uso del transporte
- Rango socioeconómico y el uso del transporte
- Tendencias en el transporte
- Proyecciones sobre el comportamiento del transporte
- etc ...

Recordando que el problema fundamental de esta investigación es entender **la relación entre el transporte público y la población de Santiago**, es decir, teniendo información sobre los usuarios del transporte y teniendo información la población (dos fuentes distintas) **¿Qué historia podemos contar?**. Respondiendo esta amplia pregunta podemos encontrarnos con patrones, con fallas, con implementaciones satisfactorias, con posibles mejoras, proyecciones, etc.



3. Datos

Para llevar a cabo este proyecto, dispondremos, principalmente, de dos fuentes de datos.

1. Informes del Directorio de Transporte Público
2. Información recopilada hasta el último censo (disponible) 2017
3. Otra fuente si se encuentra ...
4.

En el primero podremos encontrar diferentes índices, estadísticas y registros que pertenecen al **DTP**. Lo bueno de esta base, es que los datos son estructurados. Podemos encontrar las siguientes tablas de información :

-
-
-

Por otro los datos del **INE** cuentan con un abanico más amplio de opciones, pero por lo mismo es una oportunidad para poder hacer y/o filtrar los datos de manera que nos sea útiles. En este caso haremos consultas relativas a diferentes índices considerados en los *censos* para luego poder relacionarlas con la data obtenida de la fuente anterior.

Estaremos ocupando el servicio de RedAtam Webserver, que provee una solución para las consultas a datos de esta índole. Funciona como una especie de **query** que devuelve un archivo excel con la información solicitada.



4. Preguntas de Investigación

Hay muchas cosas que podríamos preguntarnos a partir de los datos que utilizaremos.

- ¿Qué relación existe entre el volumen de los sectores etarios y el consumo de transporte público?
- ¿Cómo afectó la pandemia (COVID-19) el uso de este servicio?
- ¿Cómo se relaciona el mercado automotriz con el uso del transporte?
- ¿Cómo se relaciona el estatus socio-económico de una comuna o sujeto, con el uso del servicio?
- ¿Cómo se relaciona el cambio en el valor del servicio con el uso de este ?
- ¿Cómo afectó al metro la creación de las dos nuevas líneas? ¿Se logró contener de mejor manera la afluencia?
- ¿Qué recorridos presentan más uso, qué podemos concluir de esto ?
- ¿Cómo vemos la proyección del uso del usuario respecto a este servicio? (aumenta, se mantiene, baja)



5. Diseño tentativo

En primer lugar pensamos que lo mejor es poder ordenar (pasar a csv, json y luego pandas) toda la información que podamos recolectar, de manera que a la hora de responder preguntas podamos trabajar más fácilmente con los datos que tenemos. De esta forma la información se mantendrá intacta y podremos hacer los procesos necesarios en archivos diferentes según la pregunta a responder.

Ej :

```
1  from pasajeros import cantidad_usuarios_metro
2
3  cantidad_usuarios_metro.head()
4
5..
6
7  c_2010 = cantidad_usuarios_metro[(cantidad_usuarios_metro['ano'] >= 2000) & (cantidad_usuarios_metro['ano'] <=
      2010)]
8
9  ...
10
11 print(f'Promedio de pasajeros entre 2000 y 2010 : {prom20_10}')
```

5.1. Regresión Lineal

Dado que tenemos una gran cantidad de datos historicos, pensamo que hacer un regresión lineal puede ser una buena opción para entender (y proyectar) el comportamiento del servicio entero, como de sus diferentes ramas.

....

...

—