

# **Reporte Etapa 5: Evaluation**

<b>Evaluación de resultados</b>	<b>3</b>
Evaluación de los resultados de minería de datos	3
1. Número de viajes de cada comuna a sede	3
2. Número de viajes por cada comuna	5
3. Número de viajes por cada sede	5
4. Análisis viajes internos y externos	6
5. ¿Por qué puede variar mucho el número de viajes por sede?	7
6. ¿En qué horario existe mayor número de viajes?	7
7. Comuna de origen con la mayor cantidad de viajes por sede	9
8. ¿Qué impacto generamos con estos hallazgos?	10
9. ¿Existe posibilidad de tener nuevos objetivos de negocio o se debe realizar un nuevo proyecto?	10
Modelos aprobados	11
<b>Proceso de revisión</b>	<b>11</b>
<b>Determinar los siguientes pasos</b>	<b>12</b>
Lista de posibles acciones	13
Decisión	13

## Evaluación de resultados

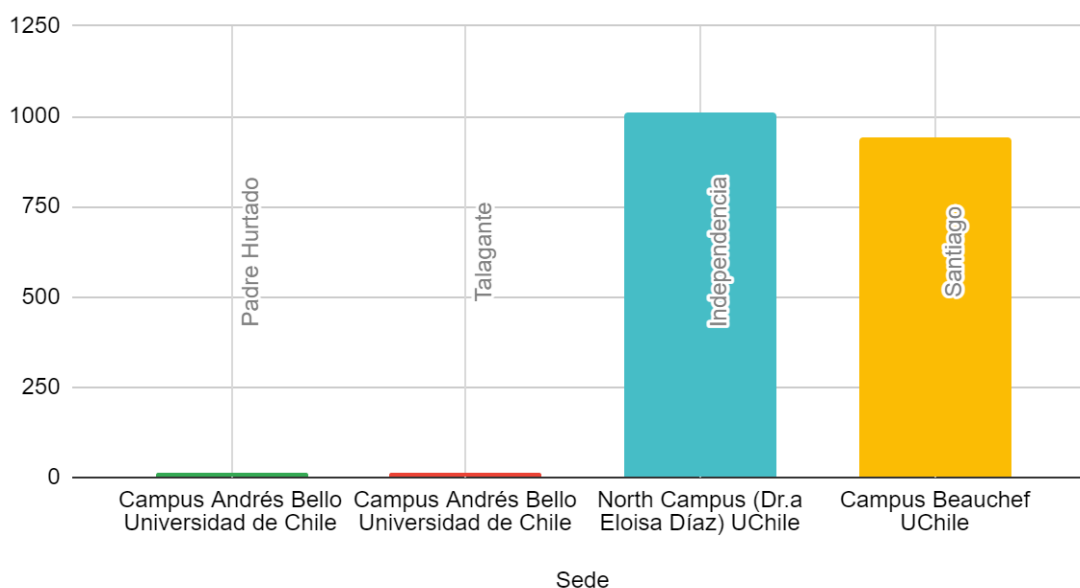
### Evaluación de los resultados de minería de datos

Los siguientes puntos de minería de datos cumplen de distintas formas los objetivos de negocio. Existen distintos puntos de vista y teorías que generan conocimiento importante para la toma de decisiones que estamos seguros que ayudarán a tomar decisiones para el Sistema de Transporte Urbano:

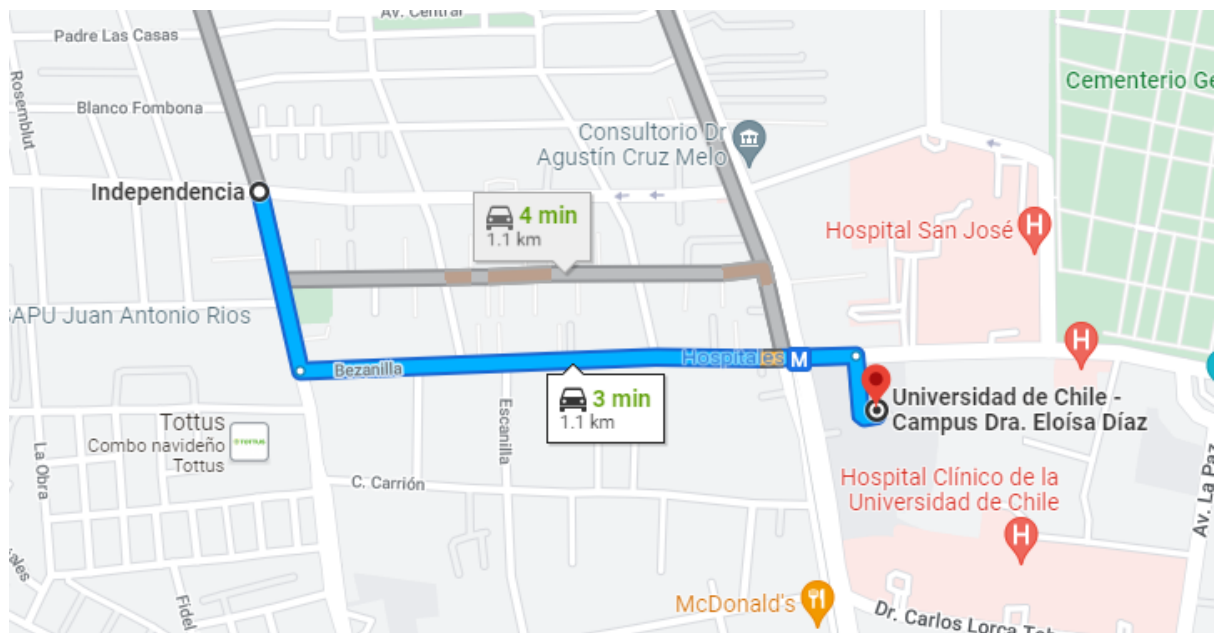
#### 1. Número de viajes de cada comuna a sede

Este resultado fue uno de los más sencillos de evaluar y comprender, ya que como esperábamos los resultados daban mucho sentido. Como por ejemplo las comunas de Padre Hurtado y Talagante hacia el Campus Andrés Bello de la Universidad de Chile tienen solo un viaje, ya que la distancia de la universidad a la comuna es aproximadamente de 40 minutos en carro. Al contrario la comuna de Independencia hacia la North Campus (Dra. Eloisa Díaz) Uchile reportan la mayor cantidad de viajes, ya que la comuna está pegada al área de la universidad. Viendo estos resultados podríamos decir que las personas eligen una zona universitaria por la cercanía.

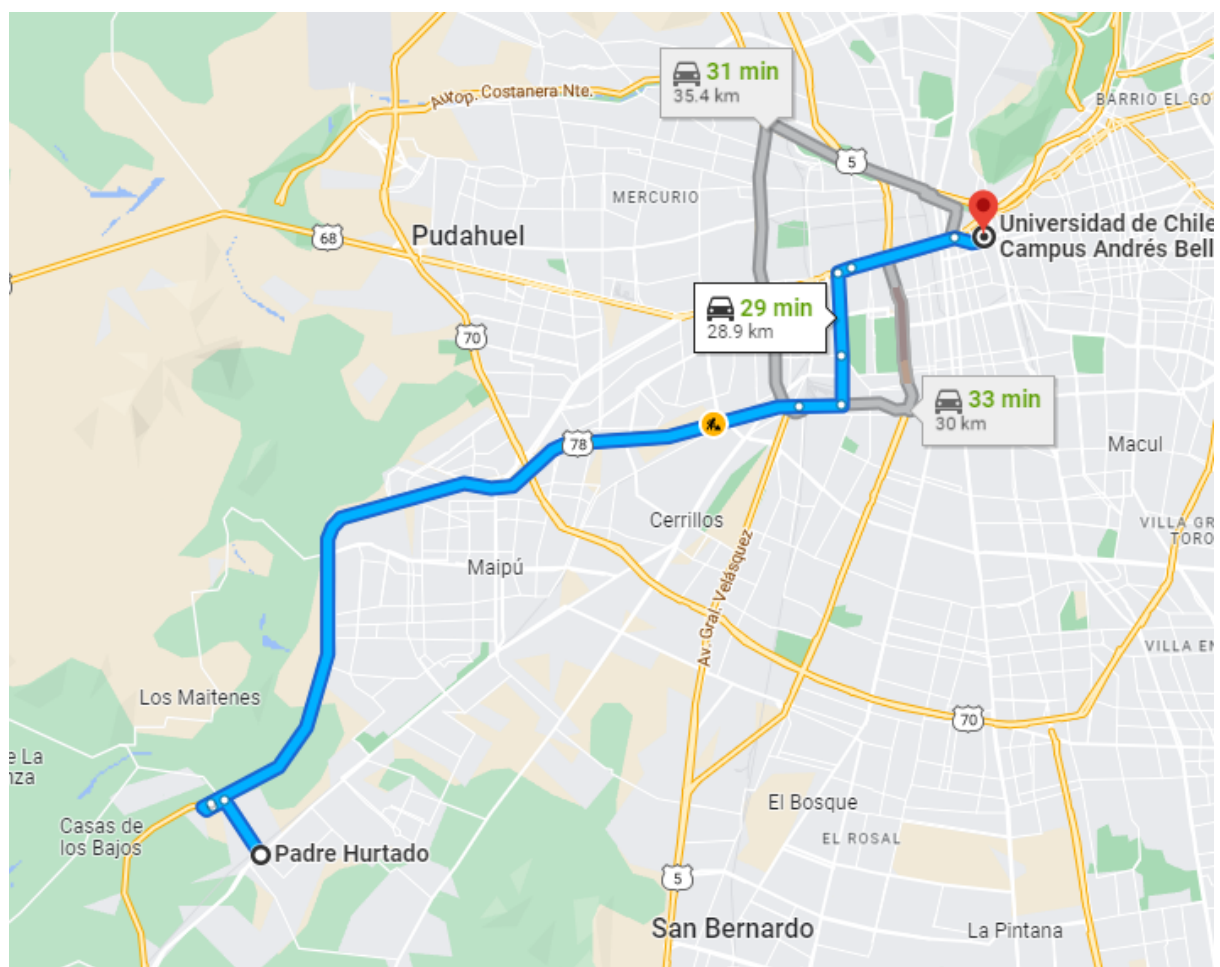
Comuna y Sede vs Viajes



Por ejemplo, una comuna con muchos viajes casi siempre tiene una distancia corta a la zona universitaria:



Y una comuna con pocos viajes, tiene una distancia muy larga a la zona universitaria:



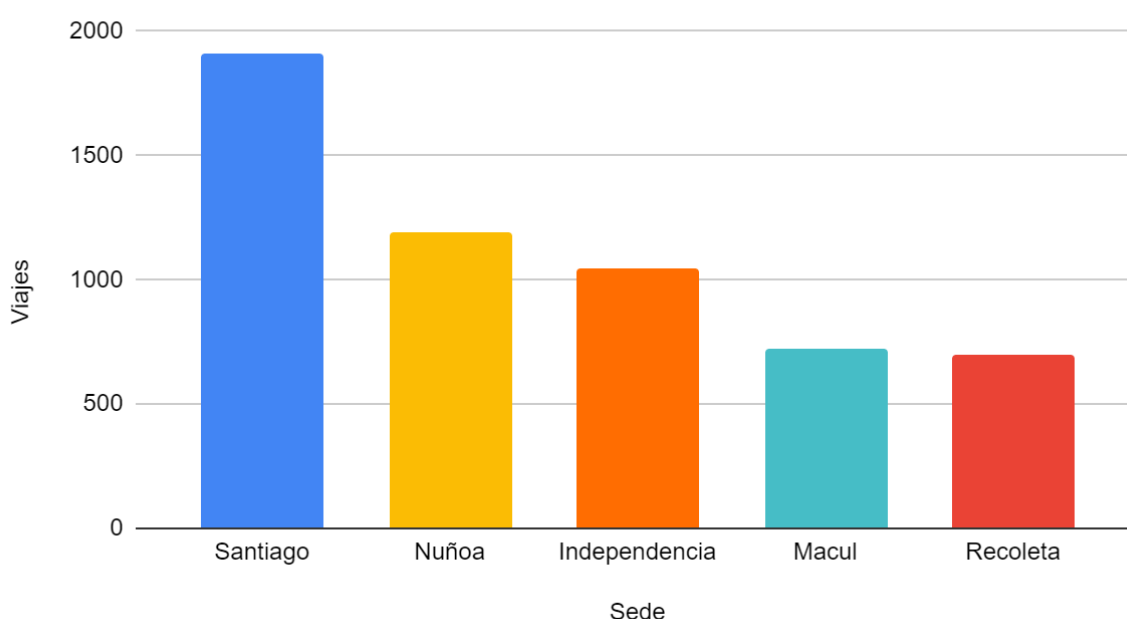
## 2. Número de viajes por cada comuna

De igual manera se nos hizo sencillo evaluar y comprender los resultados, ya que al sacar los viajes por cada comuna nos dimos cuenta que por ejemplo las primeras 5 comunas que solo cuentan con 1 a 3 viajes ya se encuentran fuera de la ciudad, lo que se nos hace obvio que no haya una gran cantidad de viajes como por ejemplo la comuna de Santiago que tiene casi 2000 viajes, esto se debe a que es el centro de la ciudad y es más fácil moverse hacia cualquier punto de la ciudad.

El top 5 de viajes por cada comuna es:

1. Santiago = 1910
2. Ñuñoa = 1189
3. Independencia = 1042
4. Macul = 721
5. Recoleta = 699

### Top viajes generados por comuna



## 3. Número de viajes por cada sede

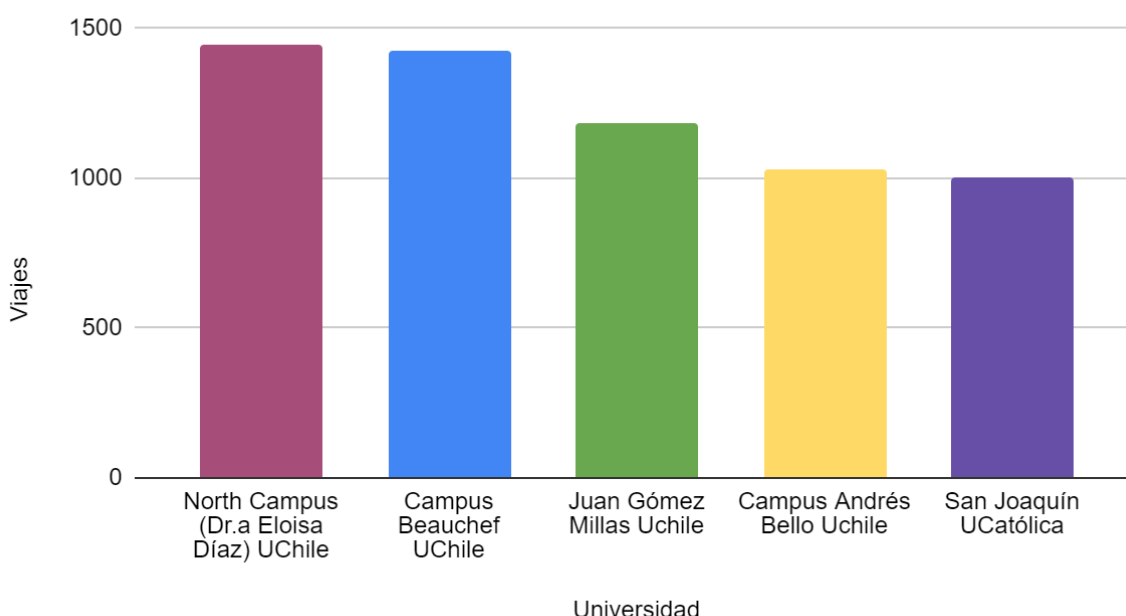
Como podemos observar en los resultados las sedes con más viajes son las de la Universidad de Chile, ya que todas las sedes se encuentran por el centro de la ciudad y sus alrededores, por lo que es más accesible para las personas y de igual manera es la única universidad pública de las que tenemos, por lo general las escuelas públicas cuentan con una mayor cantidad de estudiantes que las escuelas privadas esto incrementando la cantidad de viajes por la accesibilidad.

Excepto la Casa Central y el Campus Sur de la Universidad de Chile estas siendo las que menor cantidad de viajes cuentan de todas las sedes que tenemos, ya que el Campus Sur es la sede de la universidad que tiene una menor cantidad de carreras que las otras y la Casa Central no tiene ninguna carrera, ya que son oficinas entre otras cosas por lo que es muy poco probable que los estudiantes vayan, reduciendo mucho la cantidad de viajes.

El top 5 de viajes por cada sede es:

1. North Campus (Dra. Eloisa Díaz) Uchile = 1446
2. Campus Beauchef Uchile = 1426
3. Juan Gómez Millas Uchile = 1182
4. Campus Andrés Bello Uchile = 1027
5. San Joaquín UC = 1003

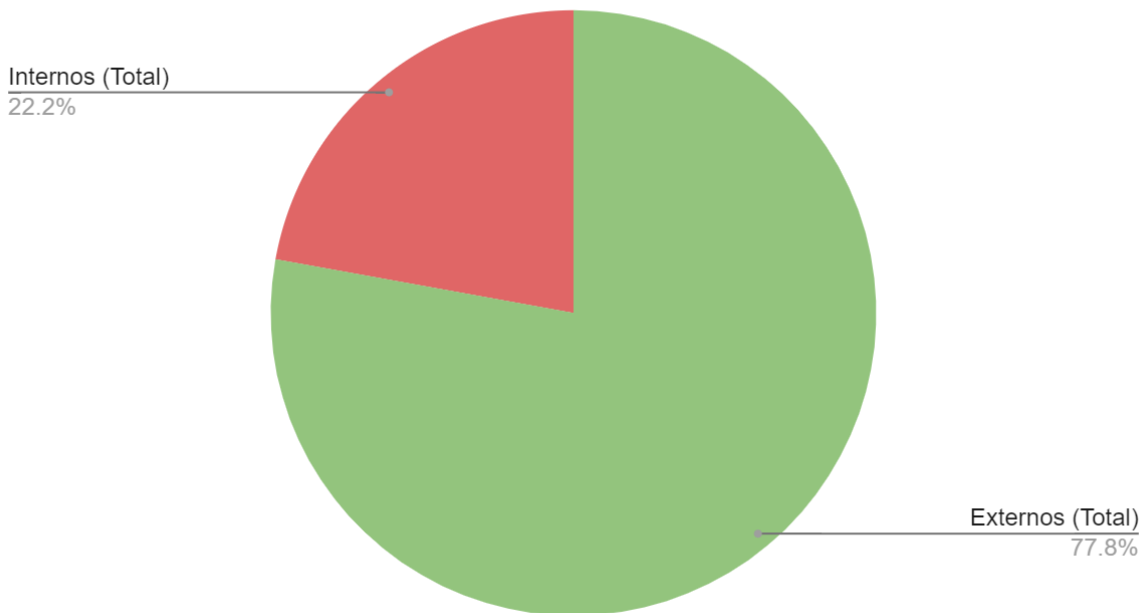
### Top 5 viajes generados por cada zona universitaria



#### 4. Análisis viajes internos y externos

Observamos que evidentemente se cuentan con más viajes externos que internos, esto porque por lo general la mayoría de personas que van a la universidad vienen de fuera. Aunque podemos ver que igual hay unos viajes internos donde pensamos que fue por la lejanía de un edificio a otro, estos viajes internos se dan más en las sedes de mayor tamaño y que cuentan con varias facultades. Se puede rectificar con la media de los viajes internos siendo de 166 y los viajes externos de 534, viéndose la clara diferencia de viajes entre uno y otro.

## Tipos de viaje

**5. ¿Por qué puede variar mucho el número de viajes por sede?**

Puede variar por muchas razones, principalmente por la cantidad de alumnos que hay en esa sede y también por la cantidad de carreras que existen en esa sede, pero también por la importancia y el tamaño de esa sede, ya sea por las facultades con las que cuenta, edificios, etc. Pero igual como vimos anteriormente en el análisis también involucra mucho la cercanía de las comunas entre las universidades. Por otro lado, teniendo en cuenta el porcentaje de la población que usa Movistar (50%), no se van a contar los viajes con 100% exactitud, así que existirán variaciones entre sedes.

**6. ¿En qué horario existe mayor número de viajes?**

La hora en la que más viajes se presentaron, en viajes internos y externos fue a la 1 de la mañana, dándonos teorías de que para los casos de viajes internos, algunos se quedan a estudiar y se regresan a sus casas que están dentro de la zona, la segunda teoría es que hubiera algún evento dentro de la zona.

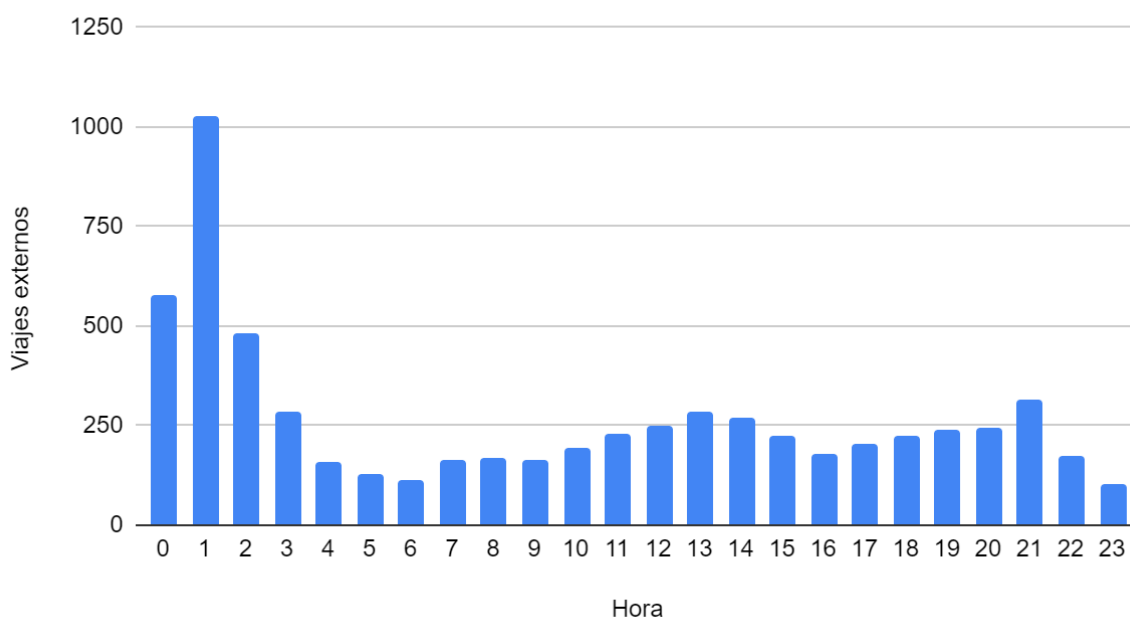
En cuanto a los viajes externos, nuestras teorías son que los estudiantes van regresado tarde a sus residencias o a sus departamentos, o que haya habido diferentes eventos nocturnos y regresan tarde de vuelta a la zona de las universidades.

Analizando estos datos con la media de viajes por hora, para el caso de viajes externos obtenemos una media de 267 y para viajes internos obtenemos una media de 76. A partir de esto buscamos los horarios que cumplan con una

cantidad mayor o superior y así podemos concluir que los rangos de hora de viajes con mucho movimiento son:

1. 12:00 am a 3:59 am
2. 01:00 pm a 02:59 pm
3. 09:00 pm a 09:59 pm

### Viajes Externos por hora

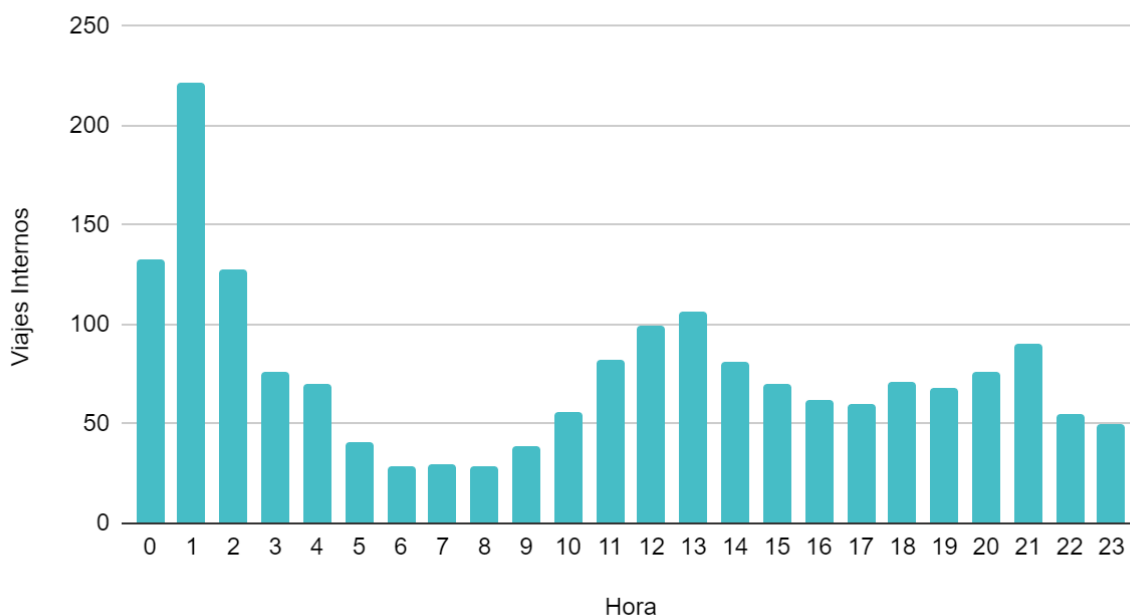


En el caso de los viajes internos, también obtenemos mucho movimiento dentro de los siguientes rangos:

1. 12:00 am a 03:59 am
2. 11:00 am a 02:59 pm
3. 08:00 pm a 09:59 pm



## Viajes Internos por hora



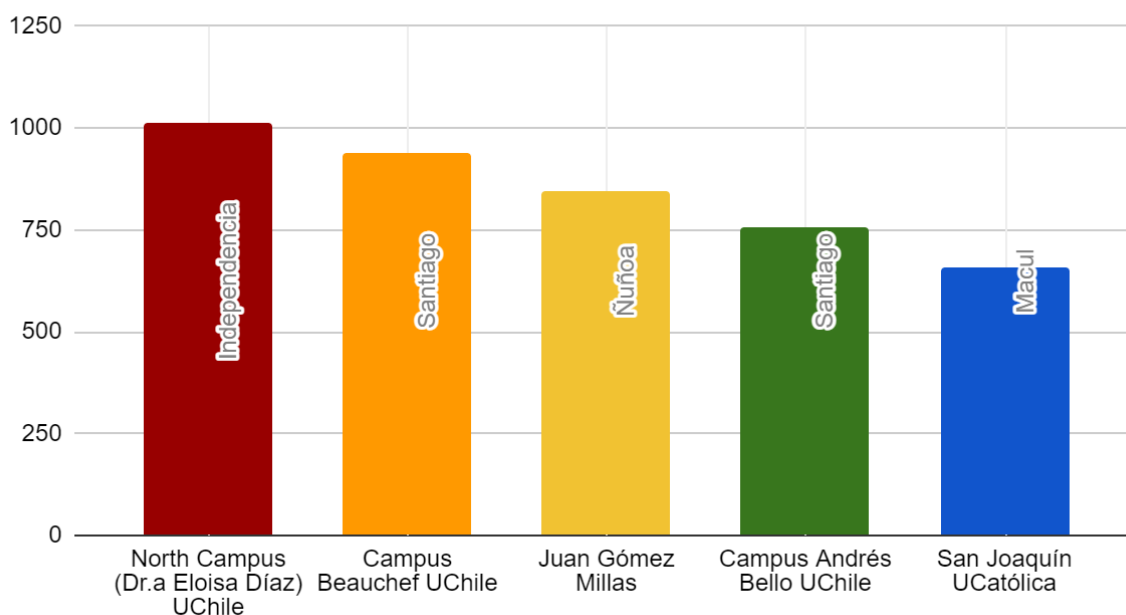
### 7. Comuna de origen con la mayor cantidad de viajes por sede

Profundizando un poco en el análisis, investigamos qué comunas generaban más viajes dependiendo la sede con más viajes. Dado los datos que obtuvimos, nos damos cuenta que son similares al número de viajes por sede, ya que aunque son una menor cantidad de viajes los resultados son casi iguales, por lo mismo que se había comentado sobre la accesibilidad que tienen esas sedes y la cantidad de carreras y alumnos con las que cuentan.

Top 5 universidades con más viajes por comuna de origen,

1. North Campus (Dra. Eloisa Díaz) Uchile  
Independencia - 1013
2. Campus Beauchef Uchile  
Santiago - 941
3. Juan Gómez Millas  
Ñuñoa - 844
4. Campus Andrés Bello Uchile  
Santiago - 759
5. San Joaquín UC  
Macul - 659

### Top 5 comunas con más viajes por zona universitaria



#### 8. ¿Qué impacto generamos con estos hallazgos?

Con los resultados obtenidos, consideramos que todo es de mucha utilidad e importancia porque estamos dando a conocer cómo es el comportamiento de la movilidad generada por las zonas universitarias de Santiago.

Con esto pudimos dar a conocer las preferencias de la población en general para asistir a cualquiera de estas zonas, como aquí se observó que es por la cercanía de los lugares y por la importancia de la zona. También conocemos los horarios en los que las personas se desplazan, las posibles razones y cómo es el desplazamiento dentro y fuera de las zonas.

Esto le permitirá a nuestra socia formadora tomar decisiones fundamentadas con el objetivo de mejorar el Sistema de Transporte Urbano.

#### 9. ¿Existe posibilidad de tener nuevos objetivos de negocio o se debe realizar un nuevo proyecto?

Consideramos que es posible continuar desarrollando este proyecto obteniendo información de más universidades con más días de viajes. Esto permitirá obtener un mejor análisis de la movilidad de las zonas durante la semana, obtener los patrones de comportamiento, predicciones confiables con uno de los modelos que hicimos y especialmente poder tomar mejores decisiones para la mejora del Sistema de Transporte Urbano.

Para futuros proyectos, creemos que la información de los datos con los que trabajamos son de mucho valor para encontrar formas de optimizar áreas muy transitadas en Santiago, considerar lugares poco transitados para nuevos

proyectos gubernamentales que mejoren la experiencia de movilidad de las personas y más cambios que permitan un buen desarrollo del Sistema de Transporte.

### Modelos aprobados

En este caso, ninguno de los modelos generados pudo ser aprobado, ya que no predicen datos lo suficientemente buenos y confiables para tomar decisiones que tengan un valor significativo en base a los objetivos de negocio, sin embargo se tuvo un modelo que demostró un mejor desempeño el cual fue el Extra Trees Regressor.

Esto quiere decir que los modelos son deficientes y no dan información de utilidad. Con los datos de predicción no podemos saber qué comunas pueden generar más viajes, no podemos saber qué universidades pueden afectar la cantidad de viajes, los desplazamientos internos entre zonas, las horas más activas del día en las zonas o comunas de origen y más.

Las predicciones no permiten proponer mejoras para el Sistema de Transporte Urbano.

### Proceso de revisión

Principalmente, se identificó una necesidad de obtener conocimiento de la movilidad de la ciudad de Santiago, Chile, enfocándonos en el efecto que tienen las zonas universitarias dentro de esta. Se realizó un proceso de preparación de datos en el que se tomó el dataset original para crear en base de trazas una matriz de viajes en la que se describe la cantidad de viajes que se realizan de una comuna a otra. Una vez que obtuvimos esta matriz, se investigó diferentes parámetros que podrían ayudar a nuestro modelo como, la cantidad de carreras, si la universidad tiene residencias, si la universidad es privada o pública, entre otras. Con estos parámetros, podemos establecer un dataset donde se identificara la hora, la universidad, la comuna de origen y la cantidad de viajes que se realizan durante un día cualquiera.

Mediante este dataset se empezaron a investigar modelos los cuales nos fueran útiles para realizar predicciones del número de viajes basado en distintas variables, se construyeron tales modelos y se evaluaron.

Tareas que serían útiles:

- Verificar constantemente los objetivos de negocios y de minería de datos para asegurarnos que todas las decisiones tomadas sean a favor de estos.
- El plan de trabajo debería de estar más relacionado a la metodología.
- Visualización de datos y Exploración de datos más profunda.

En base al objetivo se realizó una selección de modelos, los cuales fueron: regresión lineal, random forest, gradient boosted, extra trees regressor y densely connected neural network.

Para la futura evaluación se generó un diseño de pruebas para la comparación del desempeño de los modelos para determinar la calidad del cumplimiento de los objetivos de negocio. Posteriormente se construyeron los modelos, implementando así 5 modelos diferentes con sus respectivos ajustes de parámetros para la mejora de los mismos, hecho esto se concluyó que no se consiguió el desempeño esperado debido a los porcentajes de error elevados de los modelos, sin embargo, es posible decir que las predicciones del mejor modelo que obtuvimos podrían ser de ayuda al momento de tomar decisiones acerca de la movilidad.

Este proceso resultó de utilidad para determinar el éxito del proyecto en base a los objetivos definidos en la primera etapa.

El proceso de minería de datos podría ser mejorado al tener una documentación con datos más explícitos de cada proceso realizado previamente, como la descripción de unidades de tiempo, distancia, y velocidad. De esta manera así evitamos errores al procesar datos, también ayudaría mucho contar con mejores recursos computacionales o aprender a utilizar librerías más eficientes que hagan el uso de varios núcleos de procesamiento para optimizar el tiempo durante el procesamiento de datos.

Durante el proceso, se debió de realizar una mejor documentación de los supuestos y datos que eran relevantes para el procesamiento, ya que la etapa en la que más se tuvo fallas fue en la etapa de preparación de datos. Posibles acciones para corregir esto, es llevar un log donde se lleve un control de las personas que modificaron documentos y cómo se modificó para poder tener acceso mucho más rápido a las fallas. Tenemos herramientas como GitHub a la mano pero consideramos que un log de cambios por separado podría ser también una excelente herramienta de administración.

### **Determinar los siguientes pasos**

A partir de la experiencia que tuvimos durante la realización de este proyecto, opinamos que este cuenta con un gran potencial y un gran alcance, ya que con la cantidad de datos que se nos proporcionó, fue complicado obtener información o conclusiones antes de iniciar el procesamiento y minería de datos. Sabiendo que si el proyecto crece y abarca todas las universidades con sus sedes de todo Santiago de Chile y consideramos más días de la semana, podemos sacar unos mejores resultados para que en un futuro sean utilizados para la mejora del Sistema de Transporte Urbano y la movilización en la ciudad.

## Lista de posibles acciones

Posibles acciones	Razones a favor	Razones en contra
Analizar el potencial de entrega de cada resultado.	Analizar con una mayor cantidad de datos para mejores resultados y un mejor análisis de los mismos.	Realizar un proceso más largo y complicado.
Estimar el potencial de mejora del proceso actual.	Al contar con más datos la realización del modelo será mucho mejor por lo que tendrá un mejor potencial.	Podría llegar a ser un proceso tardado en lo que se encuentra una mejora del proceso.
Revisar los recursos restantes para determinar si se permiten iteraciones del proceso adicionales o si puede disponer de recursos adicionales.	Podría tomarse en consideración, en caso de que queramos incrementar el alcance del proyecto.	Hacer la búsqueda de los recursos y actualización del proceso.
Recomendar continuaciones alternativas.	Incrementar el dataset a una semana, para sacar un mejor modelo y resultados.	Actualizar toda la documentación y procesos, por el incremento de días.
Refinar el plan del proceso.	Mejorar el plan con nuevos objetivos de negocio y refinar el plan con las actualizaciones que se realicen.	Al refinar todo el proceso, este tendrá cambios que requerirán de re-documentar muchas partes..

## Decisión

1. Incrementar el dataset a una semana, para sacar un mejor modelo y resultados.
2. Mejorar el plan con nuevos objetivos de negocio y refinar el plan con las actualizaciones que se hagan.
3. Al contar con más datos la realización del modelo será mucho mejor por lo que tendrá un mejor potencial.
4. Analizar una mayor cantidad de datos para mejores resultados y un mejor análisis de los mismos.
5. Podría tomarse en consideración, en caso de que queramos incrementar el alcance del proyecto.

Incrementar el dataset a una semana, para sacar un mejor modelo y resultados. Esta decisión fue tomada por la falta de datos para el modelo, ya que solo era un día de viajes en el dataset.

Si llegamos a considerar por lo al menos una semana de viajes dentro del dataset, aparte de poder mejorar el modelo y tener mejores resultados, podríamos tener un mayor alcance del proyecto y ampliar la cantidad de universidades a analizar de 5 a todas las de la ciudad de Santiago de Chile. Al recabar más datos estos ya se podrían a empezar a considerar para una mejor toma de decisiones en la movilidad dentro de la ciudad y en el Sistema de Transporte Urbano.