

Reto B: *City Bike*

Objetivo

Realizar un dashboard interactivo que permita contestar o generar análisis de varias de las cuestiones planteadas.

Hacer un análisis de las rutas más utilizadas por los usuarios.

Pasos Generales del Procedimiento

1. Se describe lo que cada columna de cada tabla es, su tipo de dato y los valores que pueden tener.
2. Según los valores, se identifican las posibles columnas con *id* para diseñar un modelo de datos.
3. Las tres tablas son transformadas para tener el mismo número de columnas, donde cada columna tendrá el mismo rango de valores para todos los registros. Se apilan para tener una sola tabla.
4. De la tabla resultado se obtienen cinco tablas relacionadas como se muestra en el modelo de datos.
5. Las tablas son cargadas en BigQuery de Google Cloud Computer, de donde se consultan las diez estaciones más populares y las diez menos populares. Las tablas también son cargadas en Power BI para hacer el tablero de datos.
6. Se hace una matriz de confusión para contabilizar las conexiones entre estaciones, según dónde se inicia y dónde se termina un viaje.

- Los puntos 1 y 2 se describe con detalle del código en Python en el cuaderno 01_Exploración.ipynb
- El punto 3 se describe con detalle del código en Python en el cuaderno 02_Transformación.ipynb
- El punto 4 se describe con detalle del código en Python en el cuaderno 03_Construcción_BDD.ipynb
- El punto 6 se describe con detalle del código en Python en el cuaderno 04_Rutas.ipynb

Las 10 estaciones más populares en cada año

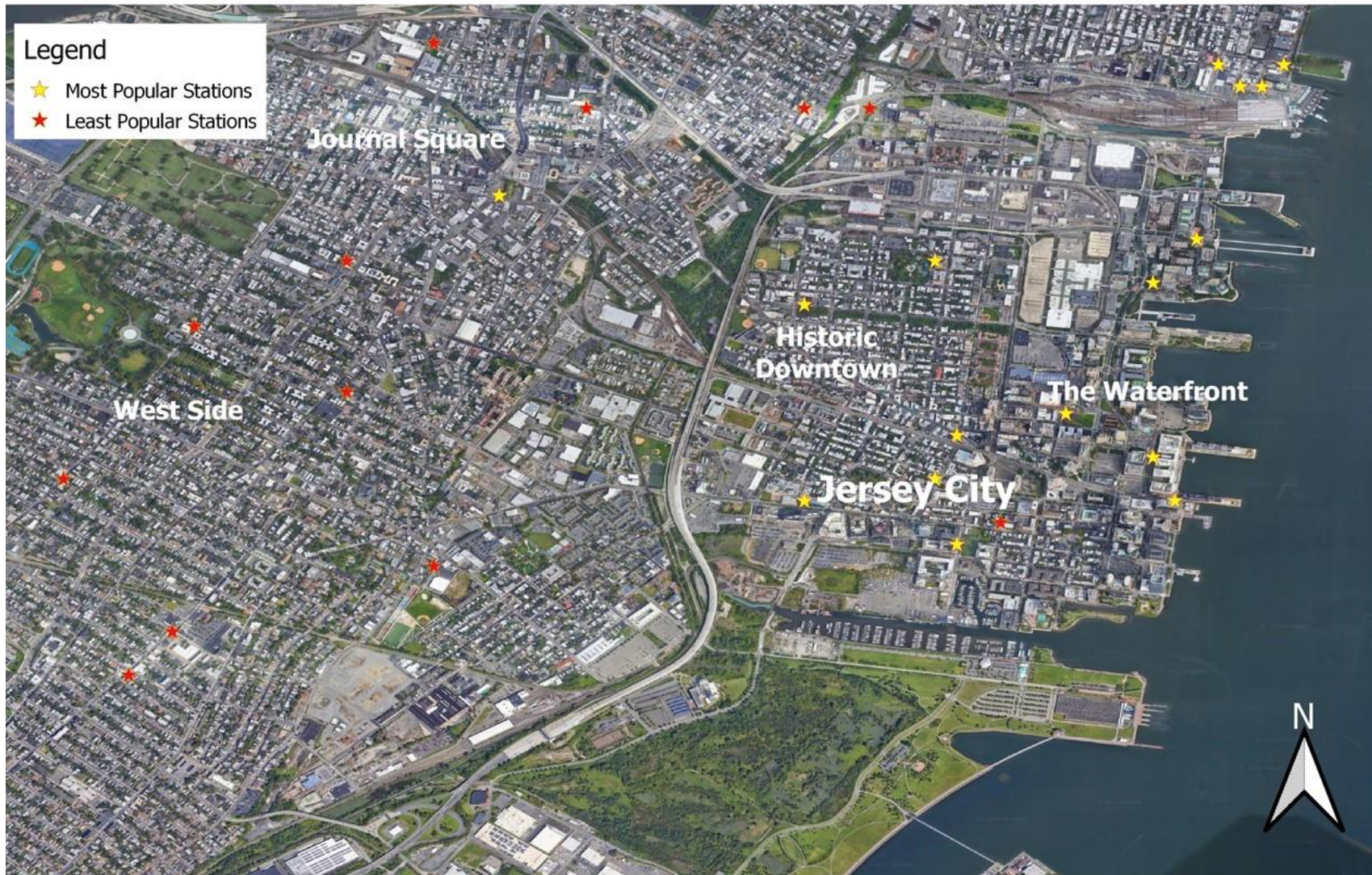
	2020		2021		2022	
Ranking	name_station	times_visited	name_station	times_visited	name_station	times_visited
1	Grove St PATH	2328	Newport Pkwy	298	Grove St PATH	1635
2	Sip Ave	1402	Grove St PATH	283	Hoboken Terminal - River St & Hudson Pl	1455
3	Hamilton Park	1202	Hamilton Park	271	Sip Ave	1316
4	Harborside	1058	Marin Light Rail	233	Hoboken Terminal - Hudson St & Hudson Pl	1060
5	Newport PATH	978	Sip Ave	229	Newport PATH	836
6	Columbus Dr at Exchange Pl	922	Newport PATH	206	South Waterfront Walkway - Sinatra Dr & 1 St	821
7	Marin Light Rail	746	Warren St	167	Newport Pkwy	801
8	City Hall	683	Harborside	164	Hamilton Park	797
9	Newport Pkwy	657	JC Medical Center	160	City Hall - Washington St & 1 St	717
10	Brunswick & 6th	654	Columbus Dr at Exchange Pl	157	8 St & Washington St	703

Las 10 estaciones menos populares en cada año

	2020		2021		2022	
Ranking	name_station	times_visited	name_station	times_visited	name_station	times_visited
1	Communipaw & Berry Lane	49	Union St	9	Grant Ave & MLK Dr	15
2	Union St	59	Jackson Square	10	Bergen Ave & Stegman St	19
3	Leonard Gordon Park	77	Heights Elevator	17	Jackson Square	25
4	Heights Elevator	99	Dey St	18	Union St	54
5	Riverview Park	99	Leonard Gordon Park	20	Communipaw & Berry Lane	64
6	Jackson Square	100	Riverview Park	22	Dey St	66
7	5 Corners Library	109	5 Corners Library	34	5 Corners Library	86
8	Dey St	114	Bergen Ave	34	Leonard Gordon Park	88
9	Glenwood Ave	120	Hoboken Ave at Monmouth St	37	Lincoln Park	137
10	Lincoln Park	134	York St	37	Christ Hospital	174

Se observa la ubicación de los dos grupos de estaciones, las más populares y las menos populares.

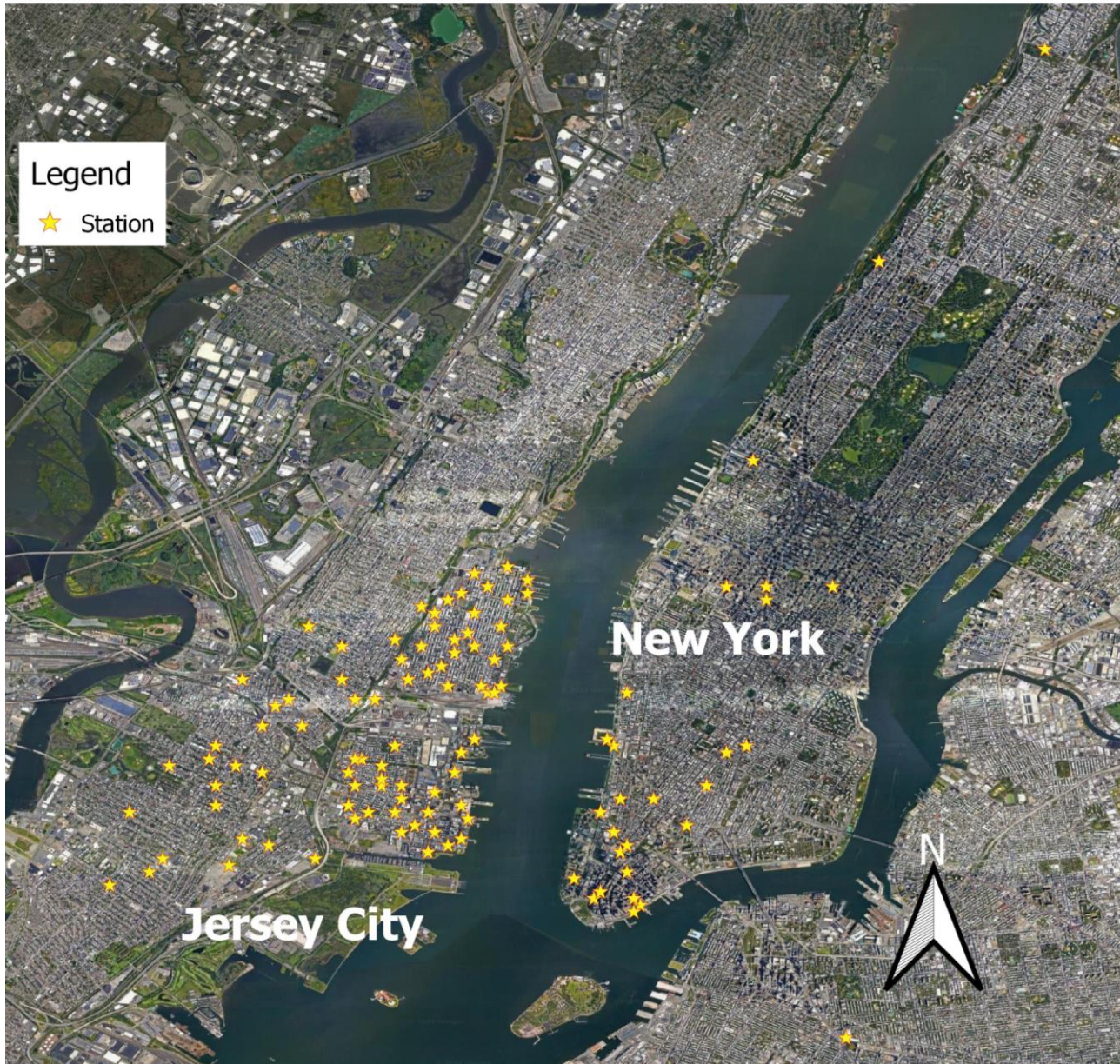
Location of the most and least popular stations over the three years



Hay una tendencia: las más populares se encuentran al este, del lado del río, mientras que las menos populares están en el oeste de la Ciudad de Jersey.

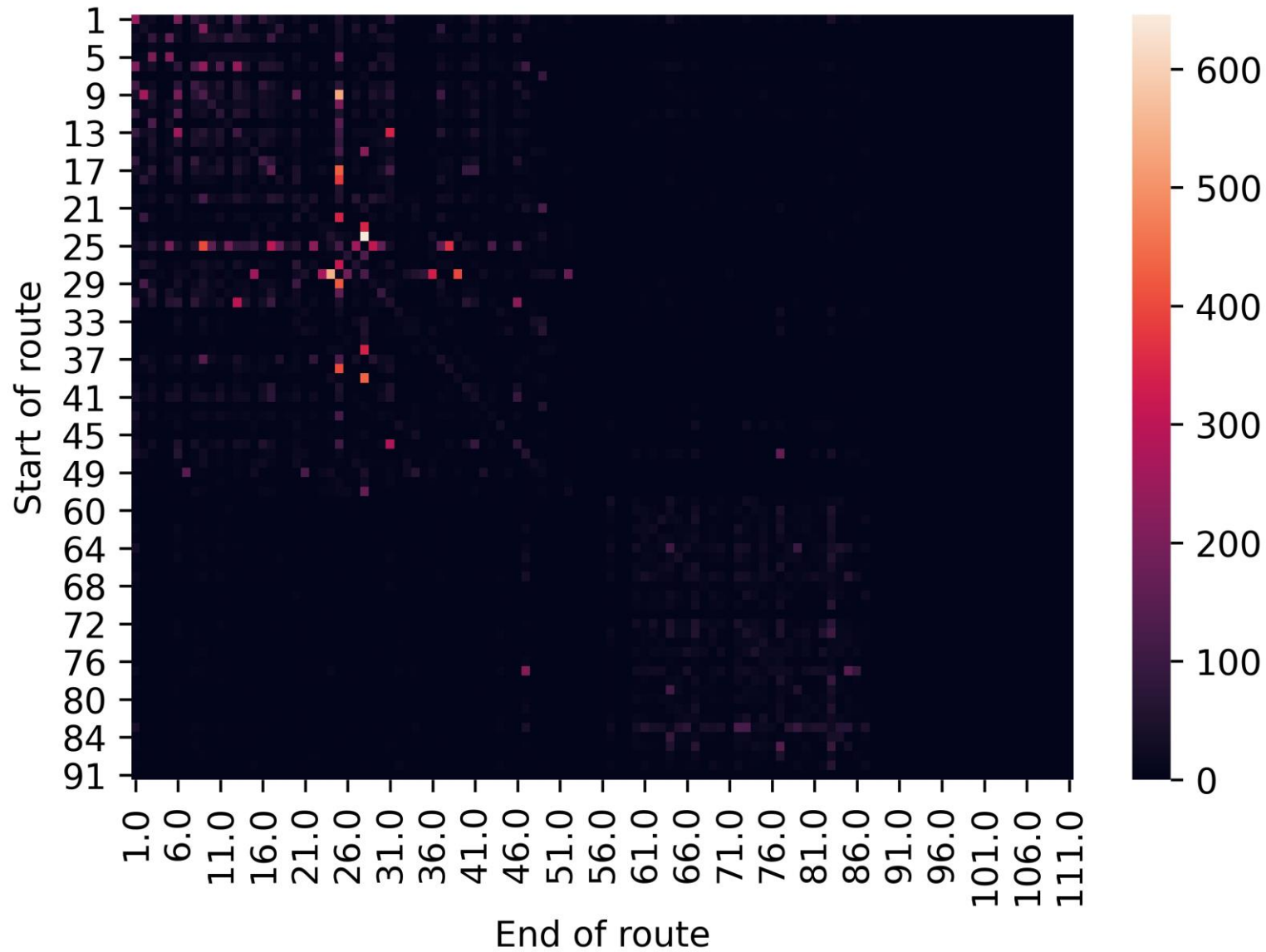
Pueden ser varios factores, como por ejemplo el atractivo visual que ofrece el río. Pero si observamos la ubicación de todas las estaciones...

Location of all stations



También hay estaciones del lado de la Ciudad de Nueva York, pero no están entre las más populares, ni las menos populares. ¿Por qué? Porque esas estaciones sólo están registradas para la devolución de la bicicleta, es decir, sólo aparecen en las columnas relacionadas con el fin del viaje.

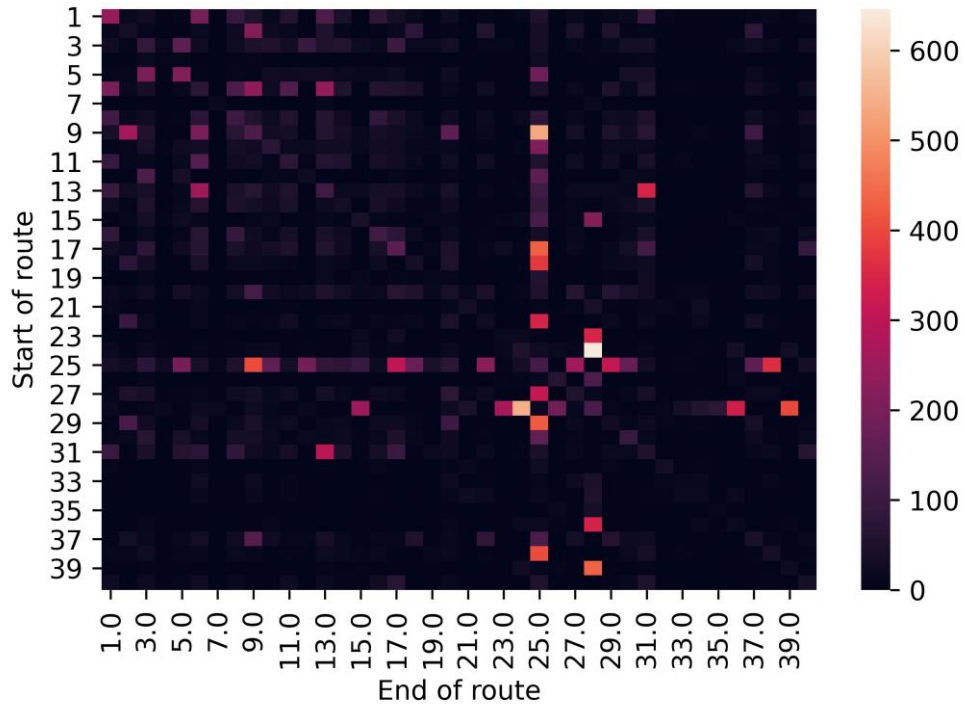
Estas observaciones sugieren hacer un análisis más detallado donde se explique qué acciones de mercado, en caso de haberlas, no han funcionado al oeste de la Ciudad de Jersey, así como la identificación de empresas del mismo giro. Y a partir de esto generar nuevas acciones a tomar para esa zona en particular.



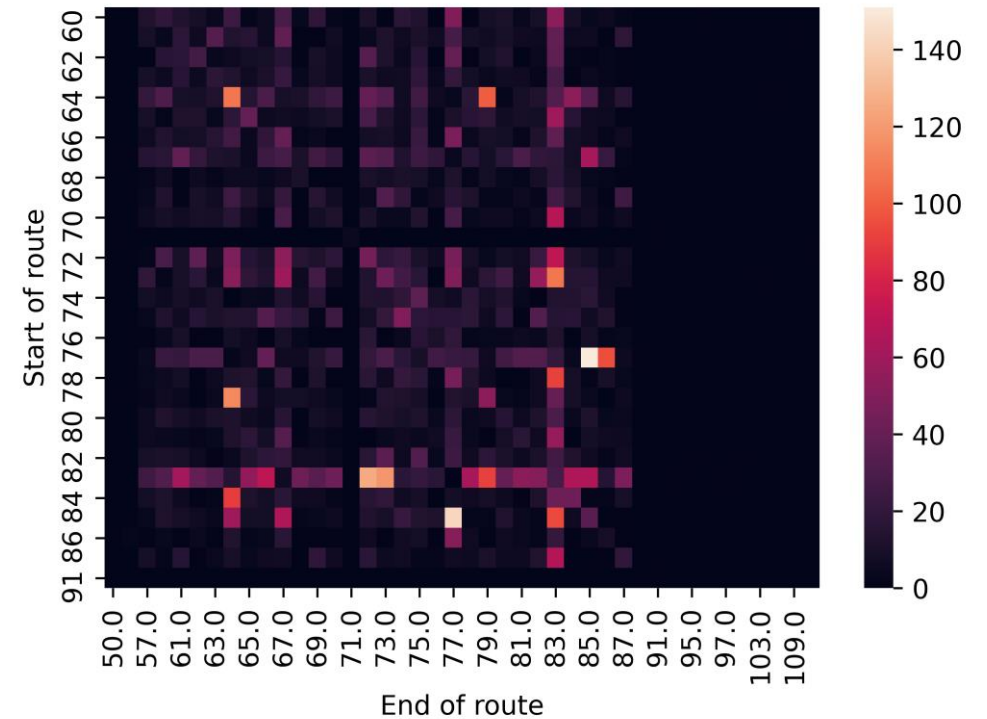
Se hace una matriz de confusión en la que se contabilizan las conexiones entre la estación donde los usuarios inician su viaje en bicicleta, hasta la estación donde lo terminan.

Dado el número de estaciones, es difícil analizar a detalle la matriz, pero se aprecian dos patrones de agrupamientos de conexiones. Uno está en la esquina superior izquierda, y otro en la esquina inferior derecha. Se hace una matriz de confusión para cada uno de esos grupos.

Urbvan

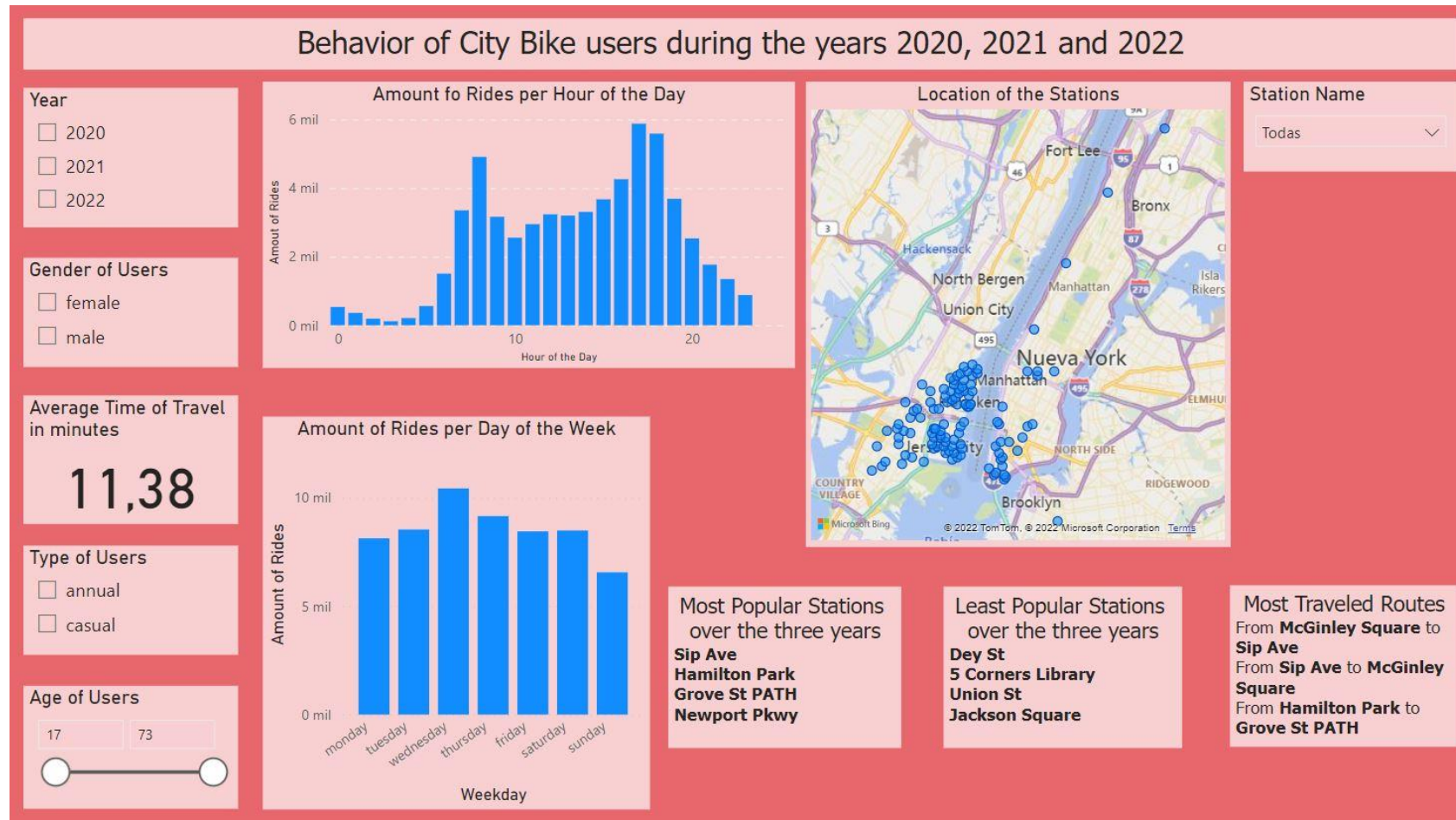


Grupo superior izquierdo: La ruta más frecuente en los usuarios es la que inicia en la estación 24 y termina en la estación 28, es decir, es la ruta *McGinley Square - Sip Ave*. Las otras dos rutas más frecuentadas son entre la estación 9 y la 25 (*Hamilton Park - Grove St PATH*). Y entre la 28 y la 24 (*Sip Ave - McGinley Square*), la cual es la ruta inversa de la de más realizada.



Grupo inferior derecho: La ruta más frecuente en los usuarios sólo para esta sección, es la que inicia en la estación 77 y termina en la estación 85, es decir, es la ruta *Hoboken Terminal, Hudson St & Hudson Pl - Madison St & 1 St*. Otra ruta notable es la iniciada en la estación 85 y finalizada en la 77, la cual de nuevo, es la ruta inversa.

Con toda la información previa (modelo de datos, rutas más frecuentes, estaciones más solicitadas, estaciones menos solicitadas) se hizo el tablero, cuya primera vista es:



Se diseñó pensando en tres secciones. A la izquierda la sección de filtros para visualizar el tablero, más un indicador del tiempo que suelen estar en la bicicleta los usuarios. Al centro los histogramas del comportamiento de los usuarios por día y hora. A la derecha la parte espacial, donde además se enlistan las estaciones y rutas ya explicadas.

Observaciones a partir de la interacción con el tablero

- En 2020 la hora preferida por los usuarios era alrededor de las 8 am, seguida por el horario alrededor de las 5 pm. En la pandemia (2021) el horario más afluente fue alrededor de las 5 pm. En 2022 Sigue siendo el horario más solicitado, pero el horario matutino, alrededor de las 8 am, comienza a mostrar una tendencia a la alza.
- Aumenta el uso del servicio entre semana conforme el usuario se acerca a la mediana edad, esto sólo con datos de 2020. Es decir, los jóvenes menores de 25 años lo usan más del miércoles al sábado. Los adultos de 26 a 50 años usan más el servicio de lunes a viernes. En cambio, personas mayores de 50 años no muestran una preferencia distintiva.
- Los usuarios que tienen membresía anual, usan más el servicio entre semana a las 8 am y a las 5 pm. En cambio, los usuarios casuales lo usan notoriamente más los sábados en horario vespertino.
- Las tres rutas más frecuentadas (McGinley Square - Sip Ave de ida y vuelta; y Hamilton Park - Grove St PATH) se dan entre semana y alrededor de las 8 am y entre 5 y 6 pm.
- Cuatro de cada 5 viajes son de personas con plan anual.

Conclusiones

- Estas observaciones permiten caracterizar el perfil del principal usuario: Adulto joven y de mediana edad con trabajo cuyo horario es de 9 am a 5 pm, aproximadamente. Prefiere trasladarse en bicicleta de forma constante, por lo que paga la anualidad del servicio.
- Durante el año pandémico disminuyó el servicio, pero no la razón 4 a 5 viajes con membresía anual, lo que sugiere la mayoría de los usuarios conservaron su plan, aunque no lo usaran. Esta hipótesis se refuerza si consideramos que en 2022 el número de viajes aumentó, pero además mantuvo una razón aproximada de 7 de cada 10 viajes fueron por un usuario con plan anual. Puedo decir que City Bike no sufrió pérdidas monetarias considerables durante la pandemia.
- En 2022 City Bike agrega nuevas estaciones, del lado de Nueva York (en 2020 y 2021 sólo tenía en Ciudad de Jersey), las cuales sólo son para finalizar el viaje, pero muestra una vez más que el negocio incluso se expandió, contrario a lo que uno pensaría como consecuencia del confinamiento.

Gracias