FACULTAD DE INGENIERIA
Universidad de Buenos Aires

# 75.06 Organización de Datos
# Primer Cuatrimestre de 2017

# Trabajo Práctico 1
# Análisis Exploratorio


| Apellido y nombre | Padrón |
|---|---|
| Palmeira, Agustín Daniel | 90.856 |
| Erramuspe, Florencia | 97.461 |
| Nicoletta, Anarella | 94.551 |
| Sánchez, Bárbara Mariana | 97.759 |

## Introducción

El objetivo del presente trabajo es realizar un análisis exploratorio del set de datos provisto por la cátedra sobre el sistema de alquiler de bicicletas público de la bahía de San Francisco.
El set cuenta con 4 archivos:  station.csv,status.csv, trips.csv y weather.csv

Trips.csv cuenta con los siguientes campos:
donde cada fila representa un viaje en bicicleta

- **id**
- **duration** (seconds)
- **start_date**
- **start_station_name**
- **start_station_id**
- **end_date**
- **end_station_name**
- **end_station_id**
- **bike_id**
- **subscription_type**
- **zip_code** (client zip code)

weather.csv cuenta con los siguientes campos:
donde cada fila tiene datos del clima para cada fecha y zipcode

- **date**
- **max_temperature_f**
- **mean_temperature_f**
- **min_temperature_f**
- **max_dew_point_f**
- **mean_dew_point_f**
- **min_dew_point_f**
- **max_humidity**
- **mean_humidity**
- **min_humidity**
- **max_sea_level_pressure_inches**
- **mean_sea_level_pressure_inches**
- **min_sea_level_pressure_inches**
- **max_visibility_miles**
- **mean_visibility_miles**
- **min_visibility_miles**
- **max_wind_Speed_mph**
- **mean_wind_speed_mph**
- **max_gust_speed_mph**
- **precipitation_inches**

- **cloud_cover**
- **events**
- **wind_dir_degrees**
- **zip_code**

Station.csv cuenta con los siguientes campos:

- **id** - unique identifier for the station
- **name** - station's name
- **lat** - latitude
- **long** - longitude
- **dock_count** - number of bikes the station can hold
- **city**
- **installation_date**

Status.csv cuenta con los siguientes campos:

- **station_id**
- **bikes_available**
- **docks_available**
- **time**

Como funciona el sistema de alquiler de bicicletas en la Bahía de San Francisco:

- Los clientes pueden optar por una membresía anual o una de 24 hs o 3 días.
- Los viajes de hasta 30 minutos son ilimitados pero en el caso de sobrepasar esta duración, se les cobra un cargo adicional, por lo tanto, los clientes deben devolver las bicicletas y pedir un nuevo código en el caso de las membresías de 24 hs o 30 minutos.

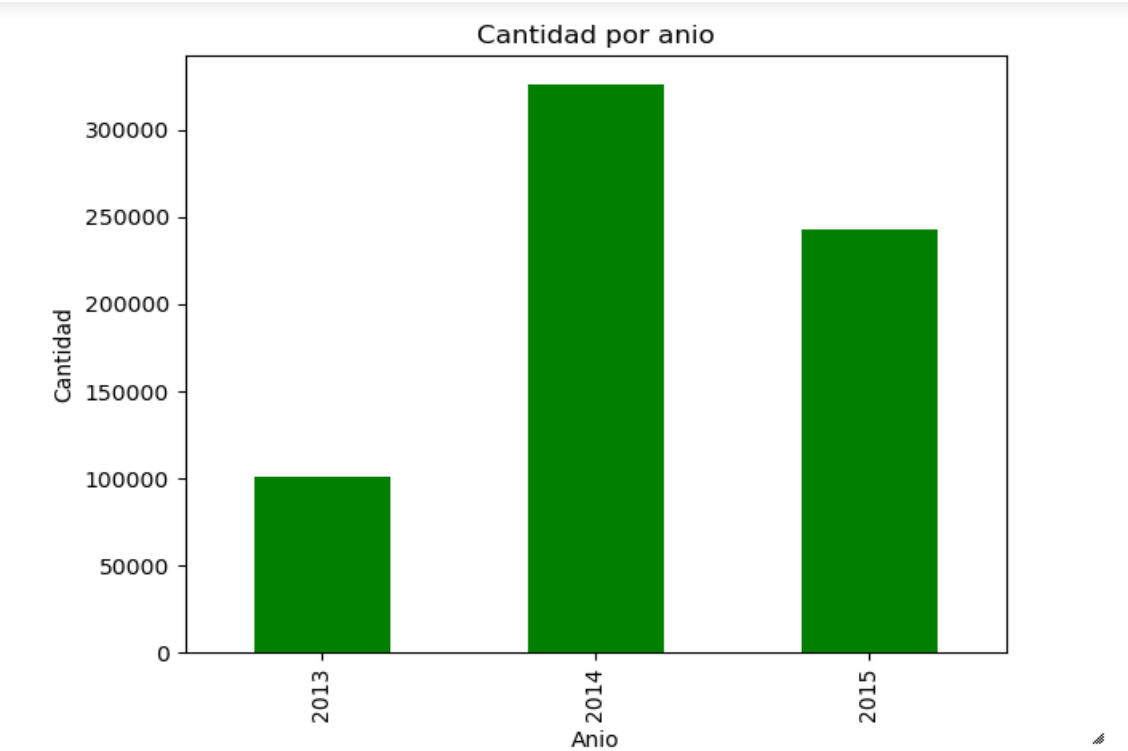
## Análisis Exploratorio

Link del repositorio:

https://github.com/agustinpalmeira/TP-BicycleRental

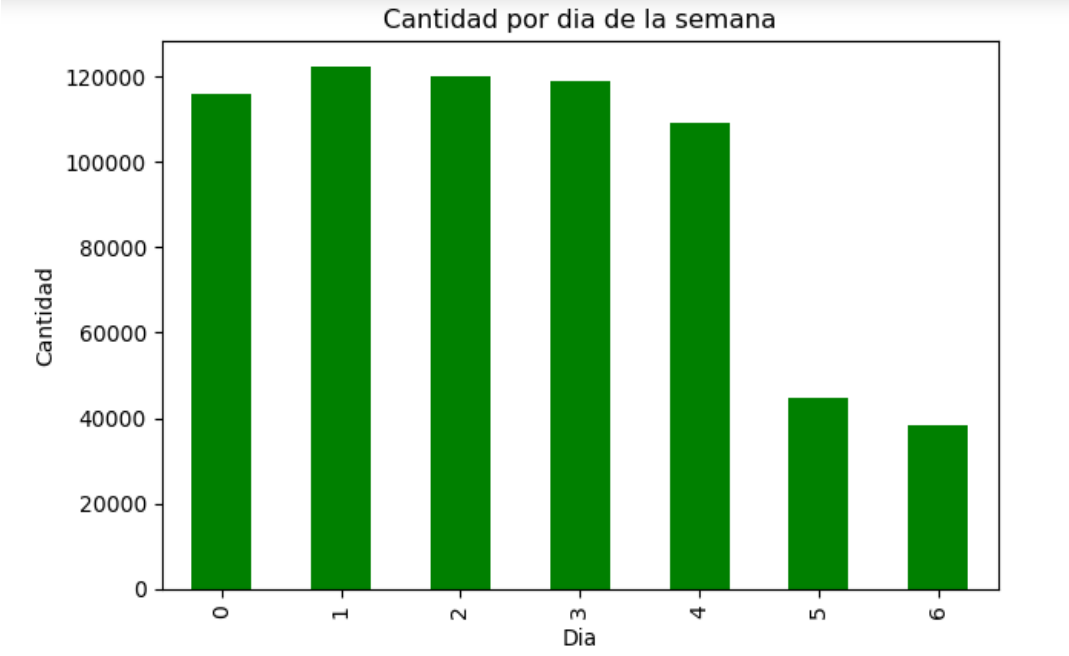
En primer lugar comenzamos analizando el archivo trips.csv y nos planteamos las siguientes preguntas:

- **¿Cómo es la distribución de la cantidad de trips por año?**




Probablemente la baja cantidad de trips registrados para el año 2013 y 2015 se deba a que trip.csv sólo registra el alquiler de bicicletas desde Agosto de 2013 hasta Agosto de 2015.

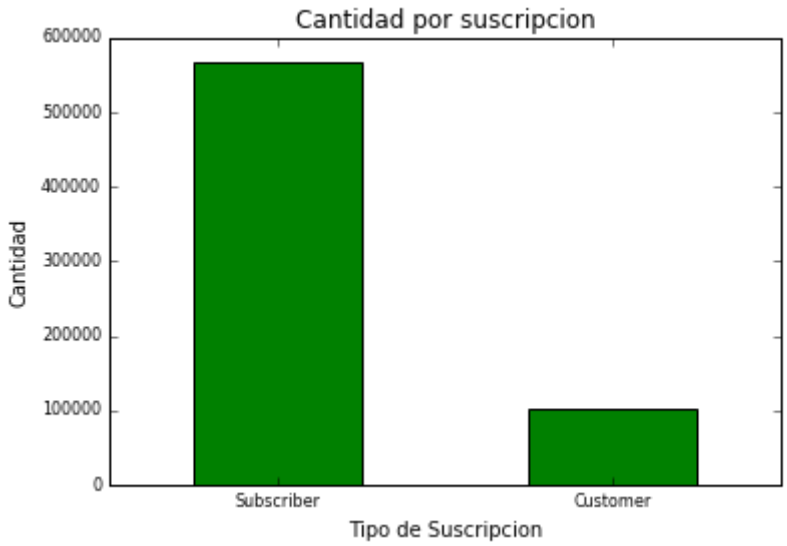
- **¿Cómo es la distribución de la cantidad de trips por día de la semana?**




0-Lunes
1-Martes
2-Miércoles
3-Jueves
4-Viernes
5-Sábado
6-Domingo

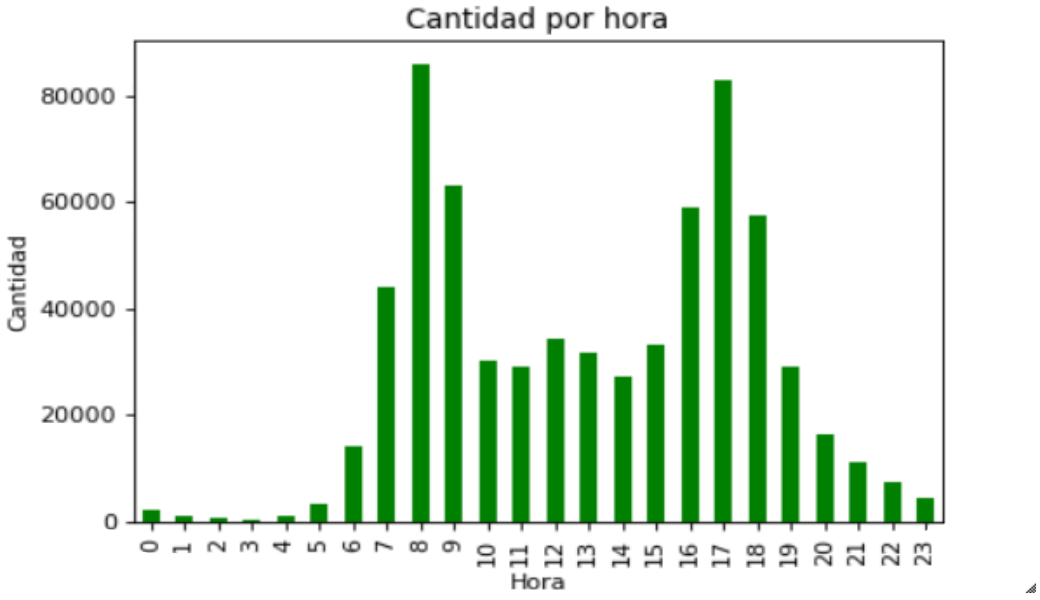
En este plot se puede visualizar que en la semana se realizaron más viajes que en el fin de semana, creemos que esto puede deberse a que la gran mayoría de los usuarios de bicicletas las utilizan como medio de transporte hasta sus respectivos empleos, escuelas y/o universidades.

- **¿Cómo es la distribución de la cantidad de trips según el tipo de membresía?**




En este plot se observa que la mayor cantidad de viajes fueron hechos por personas con una membresía anual.

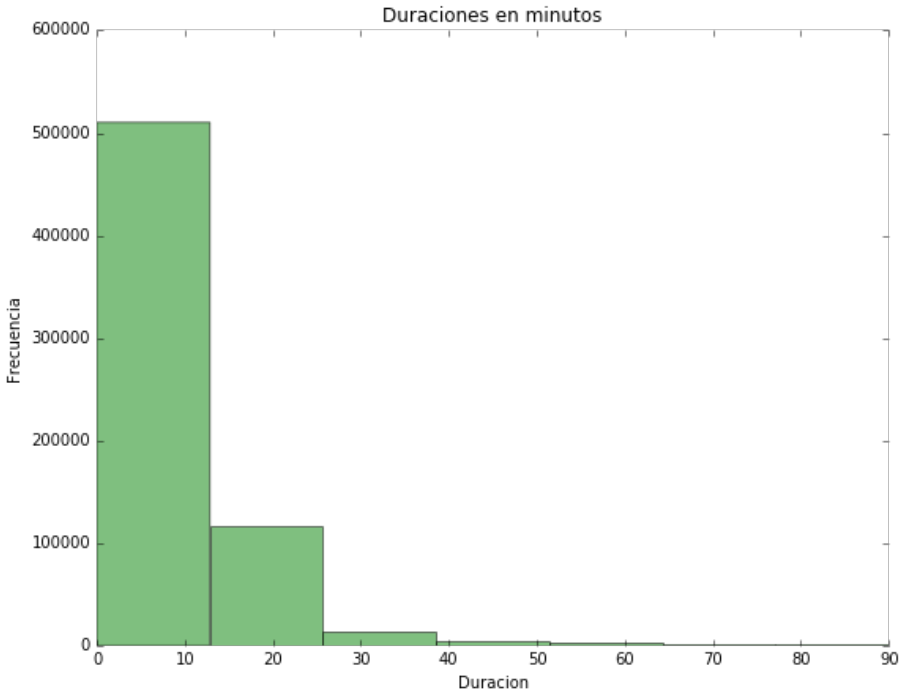
- **¿Cómo es la distribución de la cantidad de trips por hora?**




En este plot se observa que la mayor cantidad de viajes registrados fueron en los horarios: de 7 a 9 am y de 16 a 18 pm, lo que coincide con el horario pico.

Con el objetivo de llegar a conclusiones que puedan ser útiles para la empresa de alquiler, se realizaron las siguientes preguntas:

- **¿Cuál es la distribución de las duraciones de los trips en minutos?**



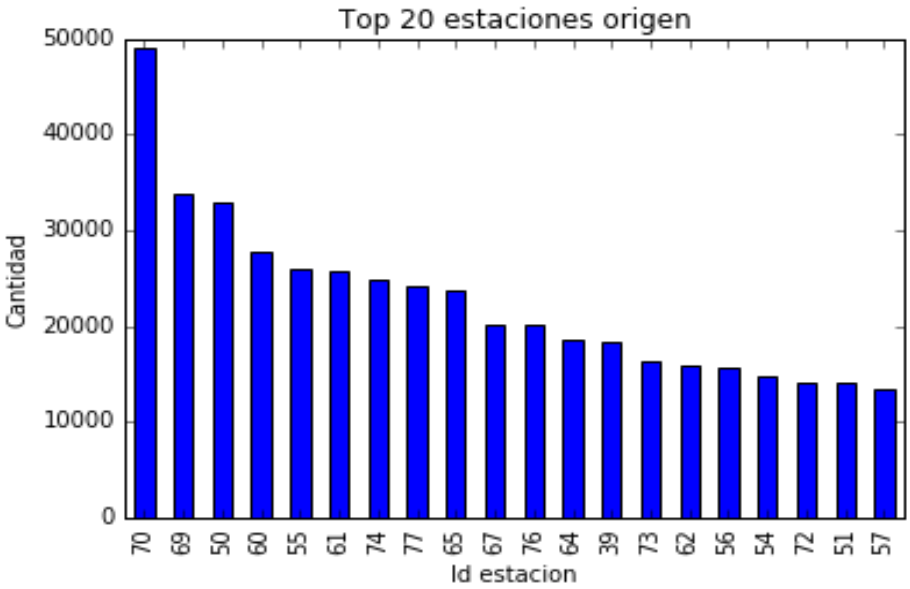

A partir del análisis realizado y de este plot, se puede afirmar que la mayor cantidad de trips fueron de una duración menor a los 40 minutos, y afortunadamente para la empresa, el porcentaje de clientes que no cumplen la regla de devolución de las bicicletas a los 30 minutos es bajo.

- **¿Cuál es el top20 del Ratio: cantidad de bicis devueltas/cantidad de bici alquiladas en una estación?**

| | |
|---|---|
| University and Emerson | 0.591270 |
| San Mateo County Center | 0.292683 |
| Redwood City Public Library | 0.253521 |
| San Jose Civic Center | 0.245295 |
| Broadway at Main | 0.238806 |
| California Ave Caltrain Station | 0.207602 |
| Palo Alto Caltrain Station | 0.187589 |
| Franklin at Maple | 0.169643 |
| Arena Green / SAP Center | 0.145722 |
| Rengstorff Avenue / California Street | 0.141718 |
| Cowper at University | 0.122999 |
| Japantown | 0.120419 |
| Park at Olive | 0.120000 |
| Stanford in Redwood City | 0.112385 |
| Evelyn Park and Ride | 0.107978 |
| Mezes Park | 0.105572 |
| Redwood City Caltrain Station | 0.099743 |
| San Jose City Hall | 0.095696 |
| San Salvador at 1st | 0.094563 |
| SJSU 4th at San Carlos | 0.094017 |

Con el objetivo de determinar en que estaciones deberían dejar más bicicletas disponibles y dónde debería haber más personal de venta, se realizó la siguiente pregunta:

- **¿Cuáles son las 20 estaciones desde las cuales parten más bicicletas?**




A partir del gráfico se puede observar que las tres estaciones desde las cuales salen más bicicletas son:

1- San Francisco Caltrain (Townsend at 4th) (id = 70), siendo el punto de partida de 49092 usuarios.
2- San Francisco Caltrain 2 (330 Townsend) (id = 69), siendo el punto de partida de 33742 usuarios.
3- Harry Bridges Plaza (Ferry Building) (id = 50), siendo el punto de partida de 32934 usuarios.

Con el objetivo de determinar las estaciones en donde debería haber más lugar para devolver las bicicletas, se planteó la siguiente pregunta:

- **¿Cuáles son las 20 estaciones a las cuales llegan más bicicletas?**

A partir del gráfico se puede observar que las tres estaciones a las cuales llegan más bicicletas son las mismas de las cuales parten la mayor cantidad de bicicletas.

1- San Francisco Caltrain (Townsend at 4th) (id = 70), siendo el punto de llegada de 63179 usuarios.
2- San Francisco Caltrain 2 (330 Townsend) (id = 69), siendo el punto de llegada de 35117 usuarios.
3- Harry Bridges Plaza (Ferry Building) (id = 50), siendo el punto de llegada de 33193 usuarios.

<u>Con el objetivo de discriminar según día de semana y fin de semana, se plantearon las siguientes preguntas:</u>

- **¿Cuáles son las 10 estaciones con más bicicletas de salida (de lunes a viernes)?**

| | |
|---|---|
| San Francisco Caltrain (Townsend at 4th) | 46234 |
| San Francisco Caltrain 2 (330 Townsend) | 31706 |
| Harry Bridges Plaza (Ferry Building) | 26520 |
| Temporary Transbay Terminal (Howard at Beale) | 25084 |
| 2nd at Townsend | 22723 |
| Steuart at Market | 22594 |
| Market at Sansome | 21932 |
| Townsend at 7th | 21655 |
| Embarcadero at Sansome | 21094 |
| Market at 10th | 17968 |

- **¿Cuáles son las 10 estaciones con menos bicicletas de salida (de lunes a viernes)?**

| | |
|---|---|
| San Jose Government Center | 22 |
| Broadway at Main | 52 |
| Franklin at Maple | 148 |
| Redwood City Public Library | 157 |
| San Mateo County Center | 258 |
| Mezes Park | 287 |
| Redwood City Medical Center | 294 |
| Stanford in Redwood City | 416 |
| Park at Olive | 574 |
| California Ave Caltrain Station | 701 |

- **¿Cuáles son las 10 estaciones más populares de salida de día laboral en horario pico(7 a 9 y 16 a 18)?**

| | |
|---|---|
| San Francisco Caltrain (Townsend at 4th) | 35048 |
| San Francisco Caltrain 2 (330 Townsend) | 23234 |
| Temporary Transbay Terminal (Howard at Beale) | 19259 |
| Harry Bridges Plaza (Ferry Building) | 18970 |
| Steuart at Market | 16249 |
| 2nd at Townsend | 15842 |
| Townsend at 7th | 14386 |
| Embarcadero at Sansome | 14137 |
| Market at Sansome | 12976 |
| Market at 10th | 11840 |

- **¿Cuáles son las 10 estaciones con más bicicletas de salida (sábado y domingo)?**

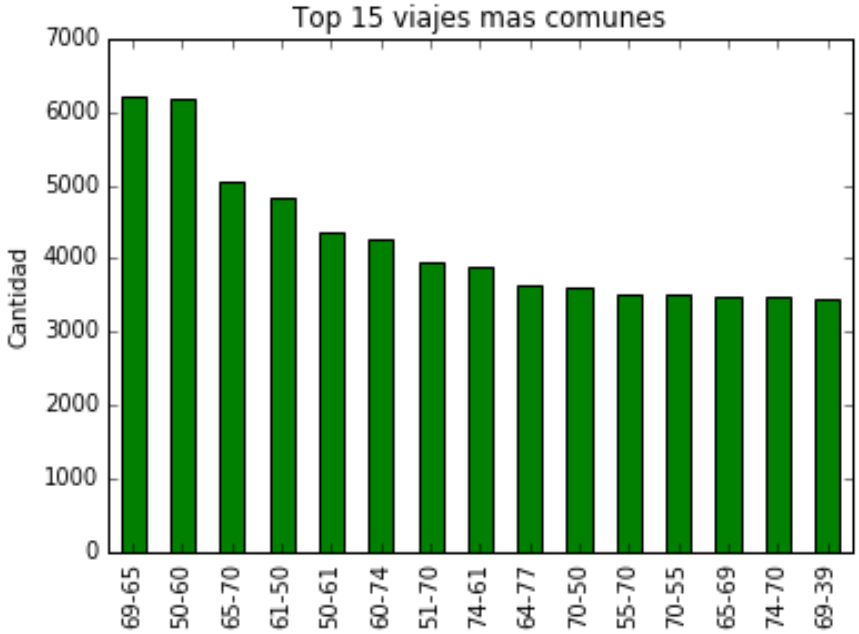
| | |
|---|---|
| Embarcadero at Sansome | 6619 |
| Harry Bridges Plaza (Ferry Building) | 6414 |
| Market at 4th | 3486 |
| Embarcadero at Bryant | 3227 |
| 2nd at Townsend | 3114 |
| Powell Street BART | 2990 |
| Grant Avenue at Columbus Avenue | 2864 |
| San Francisco Caltrain (Townsend at 4th) | 2858 |
| Powell at Post (Union Square) | 2357 |
| Market at 10th | 2304 |

- **¿Cuáles son las 10 estaciones con menos bicicletas de salida (sábado y domingo)?**

| | |
|---|---|
| San Jose Government Center | 1 |
| Broadway at Main | 15 |
| Redwood City Medical Center | 17 |
| Stanford in Redwood City | 20 |
| San Mateo County Center | 29 |
| Mezes Park | 54 |
| Redwood City Public Library | 56 |
| Santa Clara County Civic Center | 66 |
| Franklin at Maple | 76 |
| Adobe on Almaden | 115 |

Con el objetivo de determinar información útil para el Ministerio de espacio público de la Bahia de San Francisco, se decidió analizar cuáles son los viajes más comunes según origen y destino. A nuestro criterio,esta información podría ser útil para determinar cuáles son las zonas más transitadas y colocar espacios de descanso o de venta de alimentos

- **¿Cuáles son los viajes más comunes, según origen y destino?**



Se puede observar que los viajes más comunes se dan entre las siguientes estaciones, siendo origen y destino respectivamente:

1. San Francisco Caltrain 2 (330 Townsend) (id = 69) y Townsend at 7th (id=65) con una cantidad de 6189 viajes aproximadamente.
2. Harry Bridges Plaza (Ferry Building) (id = 50) y Embarcadero at Sansome (id=60) con una cantidad de 6140 viajes aproximadamente.
3. Townsend at 7th (id=65) y San Francisco Caltrain (Townsend at 4th) (id = 70) con una cantidad de 5019 viajes aproximadamente.

- **¿Cuáles son los viajes más populares de fin de semana?**

| start_station_name | end_station_name | |
|---|---|---|
| Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 1550 |
| Embarcadero at Sansome | Harry Bridges Plaza (Ferry Building) | 907 |
| | Embarcadero at Sansome | 873 |
| Harry Bridges Plaza (Ferry Building) | Harry Bridges Plaza (Ferry Building) | 841 |
| Embarcadero at Bryant | Embarcadero at Sansome | 483 |
| | Harry Bridges Plaza (Ferry Building) | 459 |
| Embarcadero at Vallejo | Embarcadero at Sansome | 451 |
| University and Emerson | University and Emerson | 448 |
| 2nd at Townsend | Harry Bridges Plaza (Ferry Building) | 415 |
| Powell Street BART | Market at 10th | 412 |

- **¿Cuáles son los 10 viajes mas populares de día de semana en horario pico(7 a 9 y 16 a 18)?**

| start_station_name | end_station_name | |
|---|---|---|
| San Francisco Caltrain 2 (330 Townsend) | Townsend at 7th | 3658 |
| 2nd at Townsend | Harry Bridges Plaza (Ferry Building) | 3264 |
| Harry Bridges Plaza (Ferry Building) | 2nd at Townsend | 3243 |
| Embarcadero at Folsom | San Francisco Caltrain (Townsend at 4th) | 3191 |
| Townsend at 7th | San Francisco Caltrain (Townsend at 4th) | 3182 |
| Embarcadero at Sansome | Steuart at Market | 3093 |
| Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 3075 |
| Steuart at Market | 2nd at Townsend | 2945 |
| Temporary Transbay Terminal (Howard at Beale) | San Francisco Caltrain (Townsend at 4th) | 2920 |
| Steuart at Market | San Francisco Caltrain (Townsend at 4th) | 2709 |

**Además realizamos las siguientes preguntas que pueden resultar de interés:**

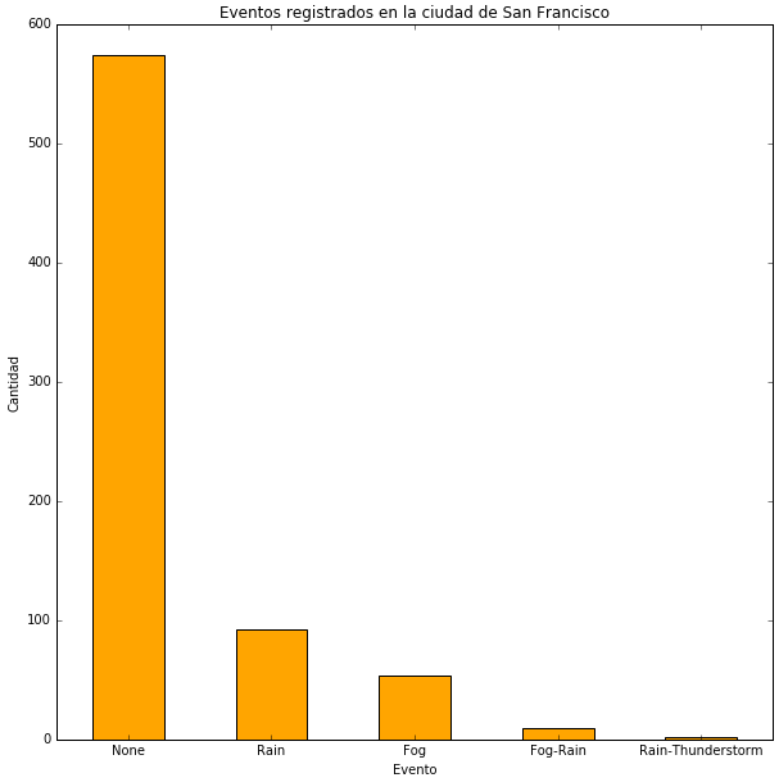
- **¿Cuál es el top10 de viajes con mayor promedio de duración, de día de semana y en horario pico?**

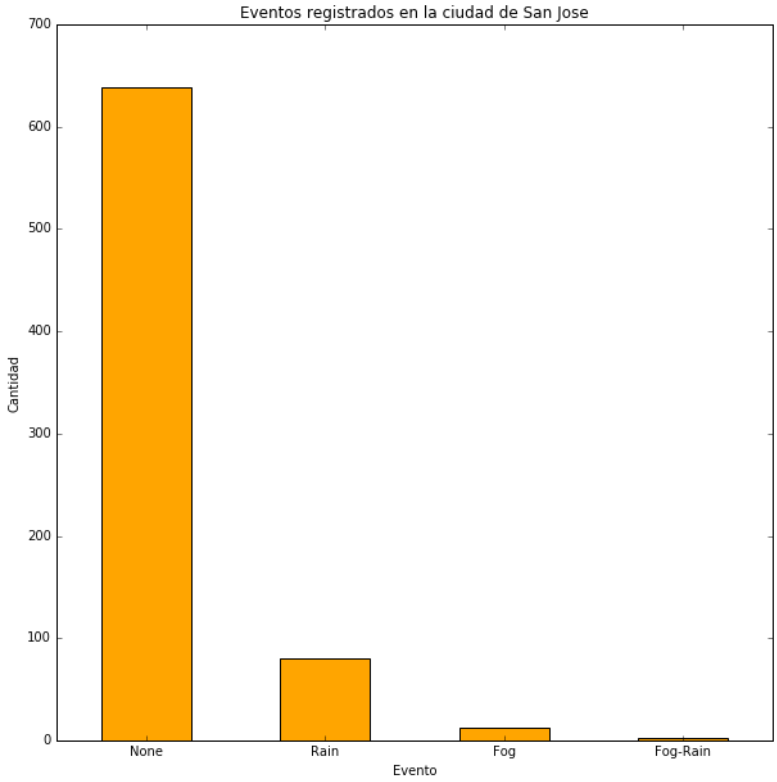
| | | size | mean |
|---|---|---|---|
| start_station_name | end_station_name | | |
| Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 3075 | 1063.781138 |
| | San Francisco Caltrain (Townsend at 4th) | 2182 | 831.148029 |
| San Francisco Caltrain (Townsend at 4th) | Market at Sansome | 2039 | 823.132908 |
| Market at 10th | San Francisco Caltrain (Townsend at 4th) | 2535 | 786.864300 |
| San Francisco Caltrain (Townsend at 4th) | Harry Bridges Plaza (Ferry Building) | 2681 | 745.986945 |
| | Temporary Transbay Terminal (Howard at Beale) | 2599 | 726.017314 |
| Steuart at Market | San Francisco Caltrain (Townsend at 4th) | 2709 | 725.410853 |
| Embarcadero at Folsom | San Francisco Caltrain (Townsend at 4th) | 3191 | 701.274835 |
| San Francisco Caltrain (Townsend at 4th) | Steuart at Market | 2357 | 699.029699 |
| Temporary Transbay Terminal (Howard at Beale) | San Francisco Caltrain (Townsend at 4th) | 2920 | 644.416438 |

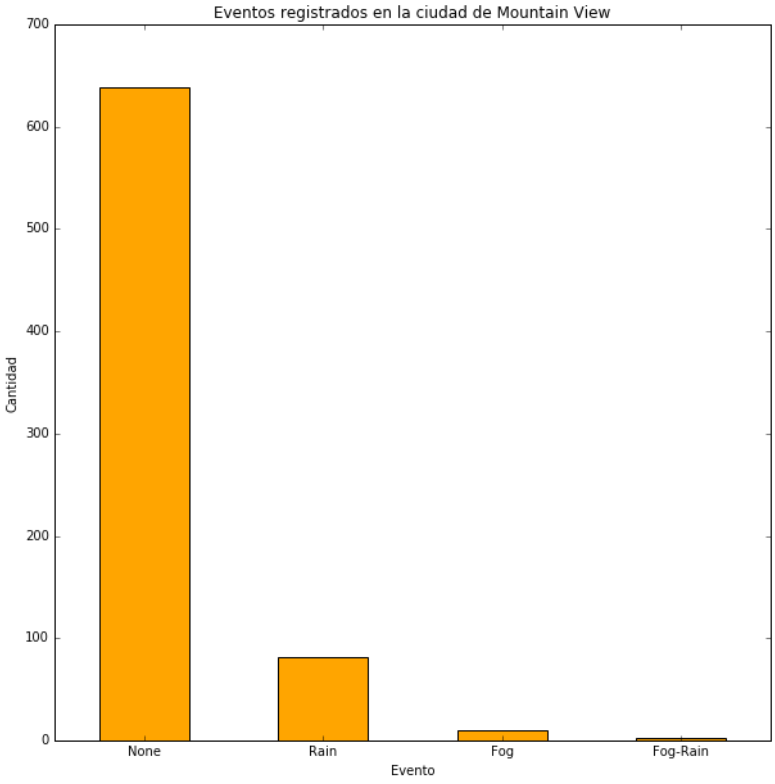
- **¿Cuál es el Top10 de viajes de día de semana y en horario pico con duración más variable?**

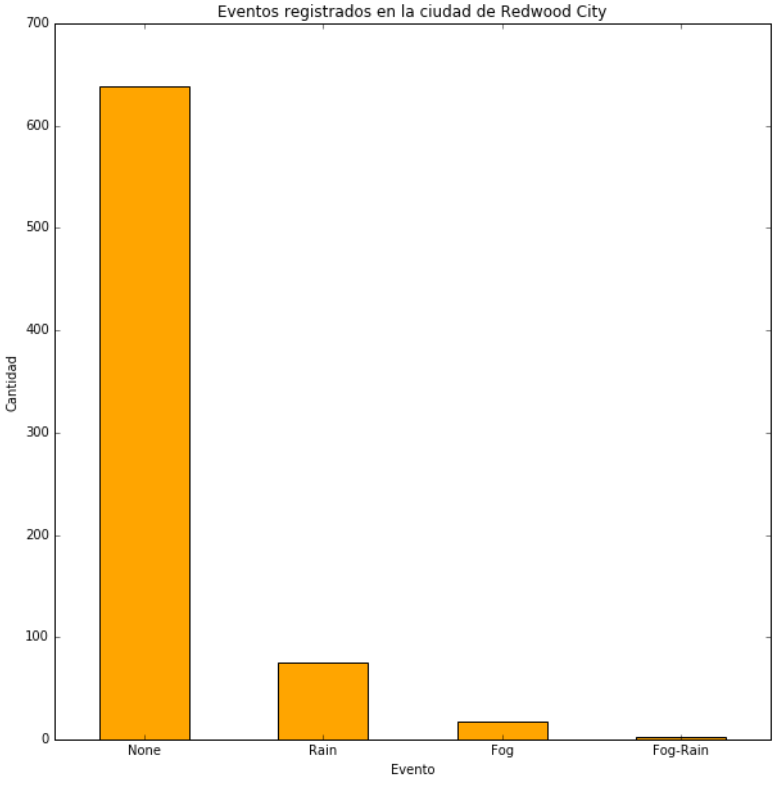
| | | size | mean | std |
|---|---|---|---|---|
| start_station_name | end_station_name | | | |
| Market at 10th | San Francisco Caltrain (Townsend at 4th) | 2535 | 786.864300 | 4523.796296 |
| Market at Sansome | 2nd at South Park | 2262 | 443.255526 | 2880.419426 |
| Mountain View Caltrain Station | Mountain View City Hall | 2085 | 411.705995 | 2692.106160 |
| Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 3075 | 1063.781138 | 2513.807266 |
| San Francisco Caltrain (Townsend at 4th) | Market at Sansome | 2039 | 823.132908 | 2452.355243 |
| Embarcadero at Folsom | San Francisco Caltrain (Townsend at 4th) | 3191 | 701.274835 | 2440.397087 |
| Steuart at Market | Embarcadero at Sansome | 2025 | 632.006420 | 1707.484570 |
| San Francisco Caltrain (Townsend at 4th) | Temporary Transbay Terminal (Howard at Beale) | 2599 | 726.017314 | 1632.042390 |
| Steuart at Market | San Francisco Caltrain (Townsend at 4th) | 2709 | 725.410853 | 1512.311844 |
| 2nd at Townsend | Harry Bridges Plaza (Ferry Building) | 3264 | 554.430453 | 1346.866430 |

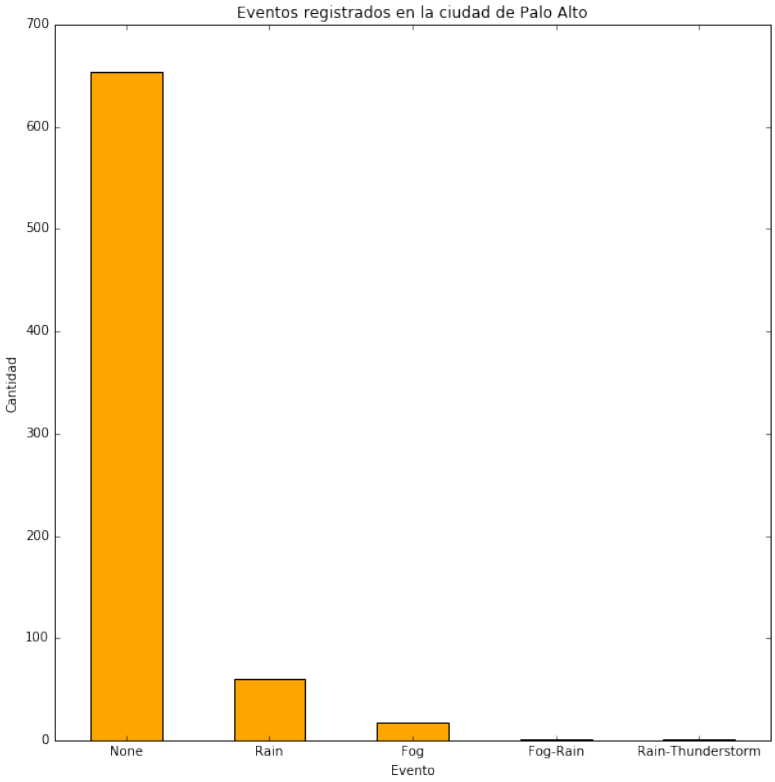
- **¿Cuáles fueron los eventos climáticos registrado para cada ciudad?**



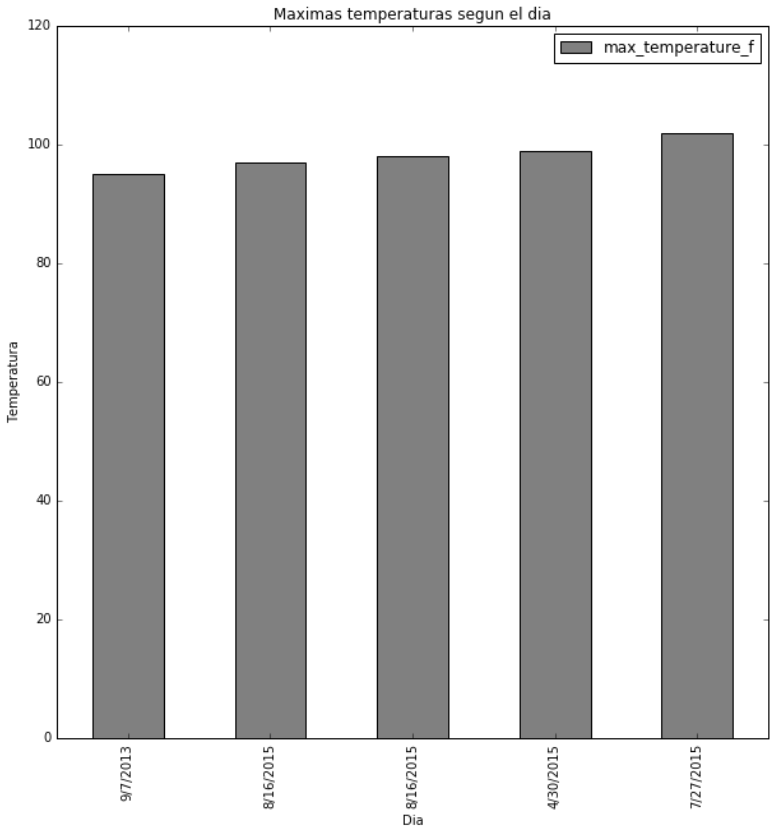



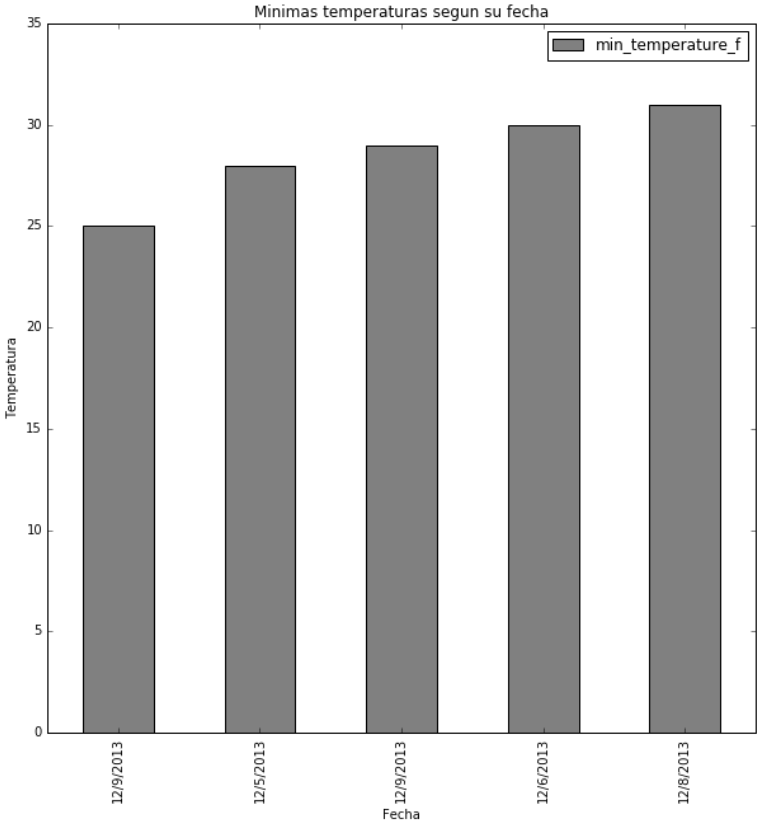

Eventos registrados en la ciudad de Mountain View
700
600
500
400
300
200
100
0
Cantidad
None
Rain
Fog
Fog-Rain
Evento

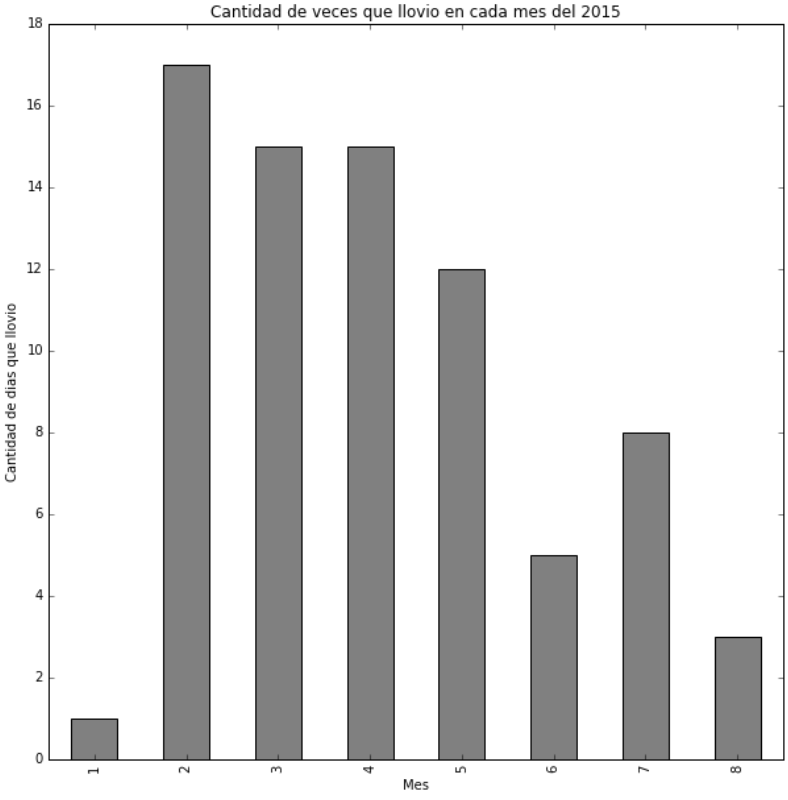

Eventos registrados en la ciudad de Redwood City
700
600
500
400
300
200
100
0
Cantidad
None
Rain
Fog
Fog-Rain
Evento

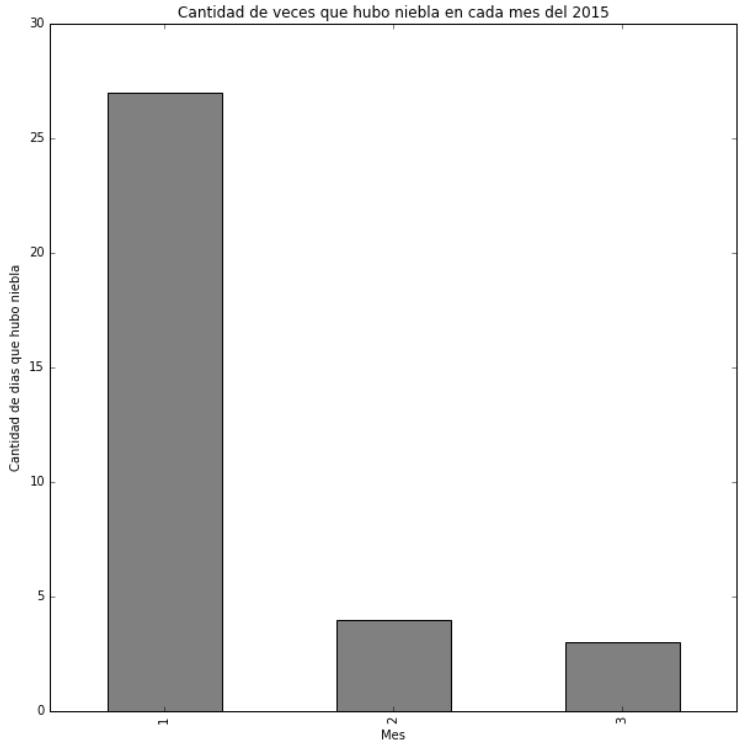



- **<u>Algunos aspectos que analizamos del clima</u>**




Minimas temperaturas segun su fecha
min_temperature_f
Temperatura
Fecha
0
5
10
15
20
25
30
35
12/9/2013
12/5/2013
12/9/2013
12/6/2013
12/8/2013


Cantidad de veces que llovio en cada mes del 2015
Cantidad de dias que llovio
Mes
0
2
4
6
8
10
12
14
16
18
1
2
3
4
5
6
7
8




**Aclaración:**

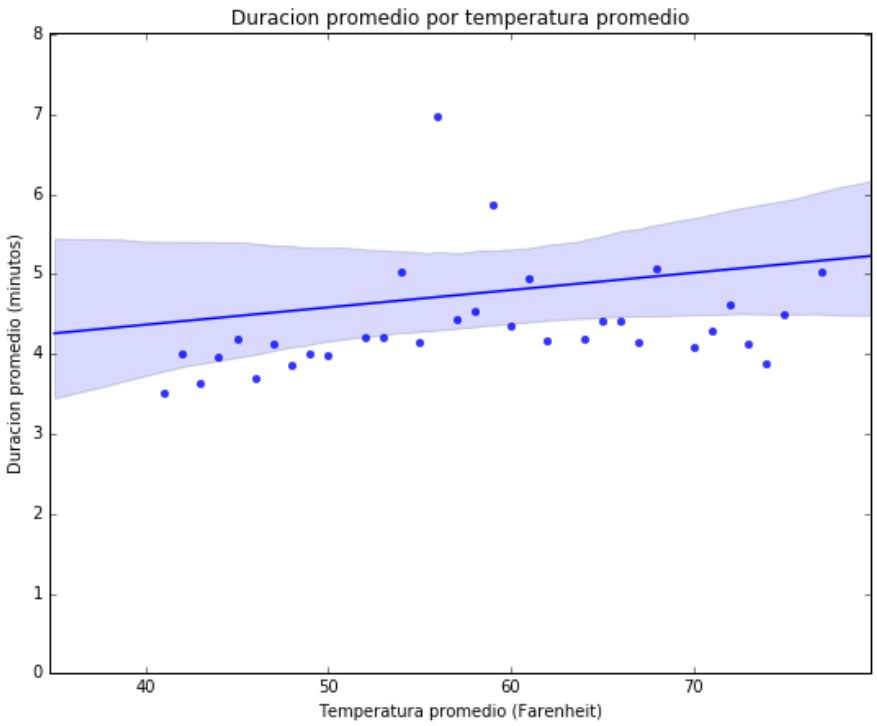
Para analizar la relación entre el clima y los trips, se hizo previamente una filtración de aquellos trips cuya duración es menor a un día, puesto que no tiene sentido analizar la relación entre las condiciones climáticas y las características del trip en los otros casos.

A su vez, también hemos considerado que las condiciones climáticas no varían notablemente entre las cinco ciudades de las cuales se tienen registro, por lo tanto, para el análisis se ha tomado el zip_code para el cual la cantidad total de valores null es la mínima (zip_code = 94107).

Nos propusimos analizar la correlación entre las distintas variables climáticas registradas y la cantidad y duración promedio de los viajes, nos hicimos las siguientes preguntas:

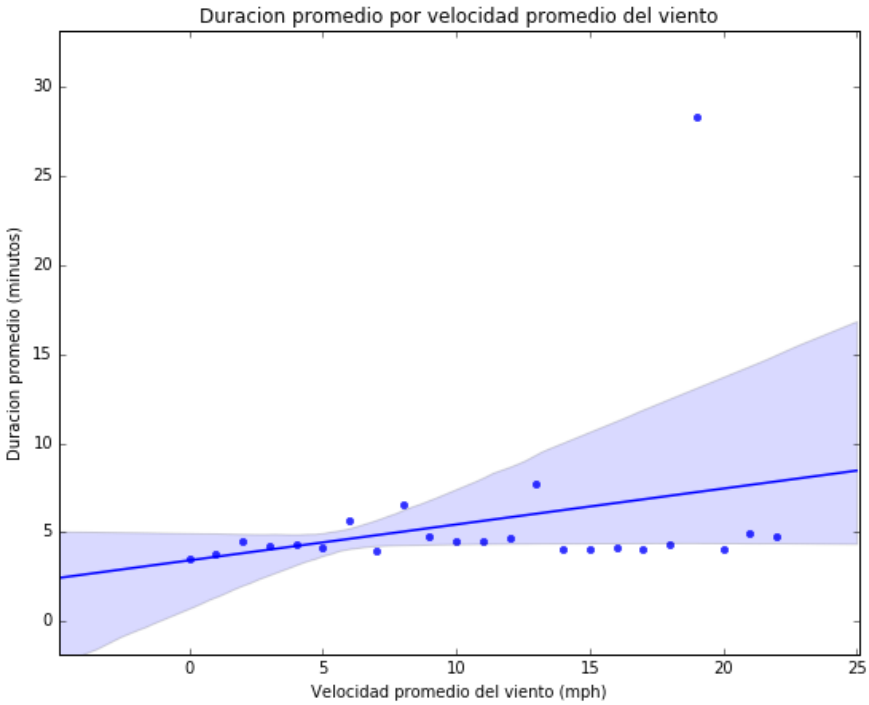
- **¿Hay correlación entre la duración promedio en minutos y la temperatura promedio en Fahrenheit para el trip más popular?**

  A partir de la siguiente visualización podemos deducir que, en general, la gente tiende a disminuir su velocidad (y por ende aumentar la duración del paseo) cuando las temperaturas son más cálidas. Esto era lo que esperábamos, puesto que la mayoría de los usuarios disfrutan más de andar en bicicleta con climas cálidos.



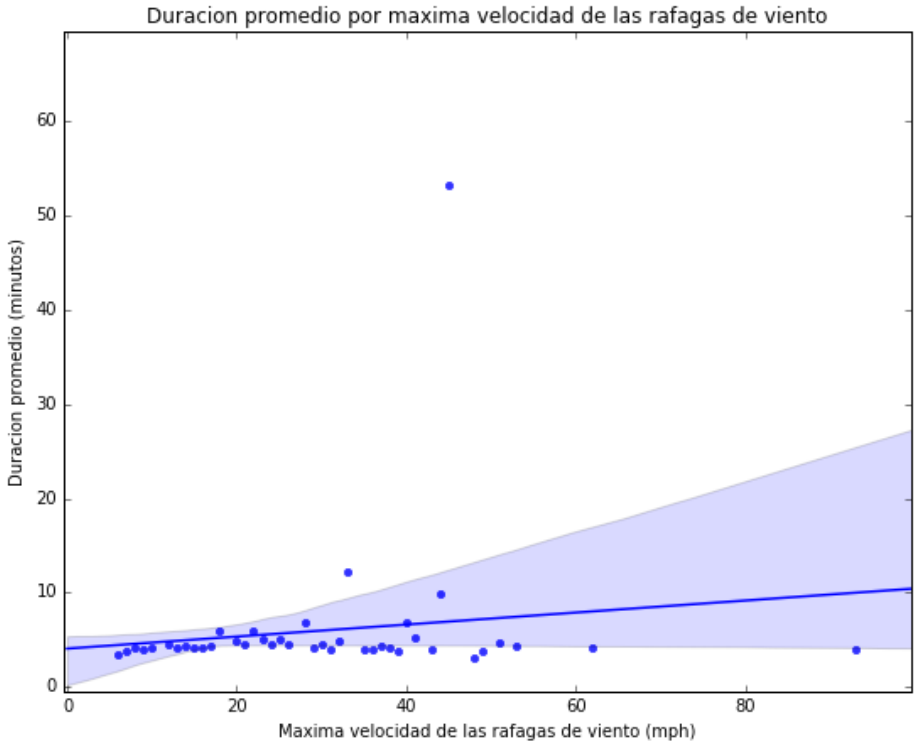

- **¿Hay alguna correlación entre la velocidad promedio del viento y la duración promedio del trip más popular ?**

  A partir de la siguiente visualización, podemos ver que en general, los paseos en bicicleta tienden a extenderse a medida que aumenta la velocidad del viento. Esto se debe, probablemente, a la resistencia que opone el viento si la persona se encuentra viajando en sentido opuesto a él.



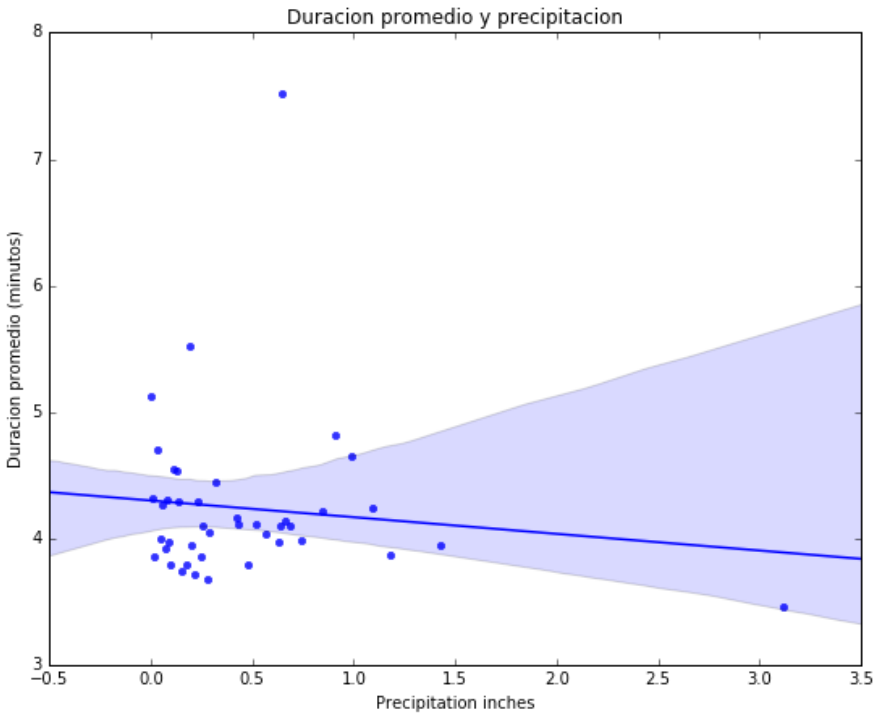
- **¿Hay alguna correlación entre la velocidad máxima de las ráfagas de viento y la duración promedio en minutos para el trip mas popular?**

  A partir de la siguiente visualización, podemos ver que, en general, la duración de los paseos suele aumentar mientras mayor es la máxima velocidad de las ráfagas de viento. Creemos que esto está directamente relacionado con lo concluido anteriormente.

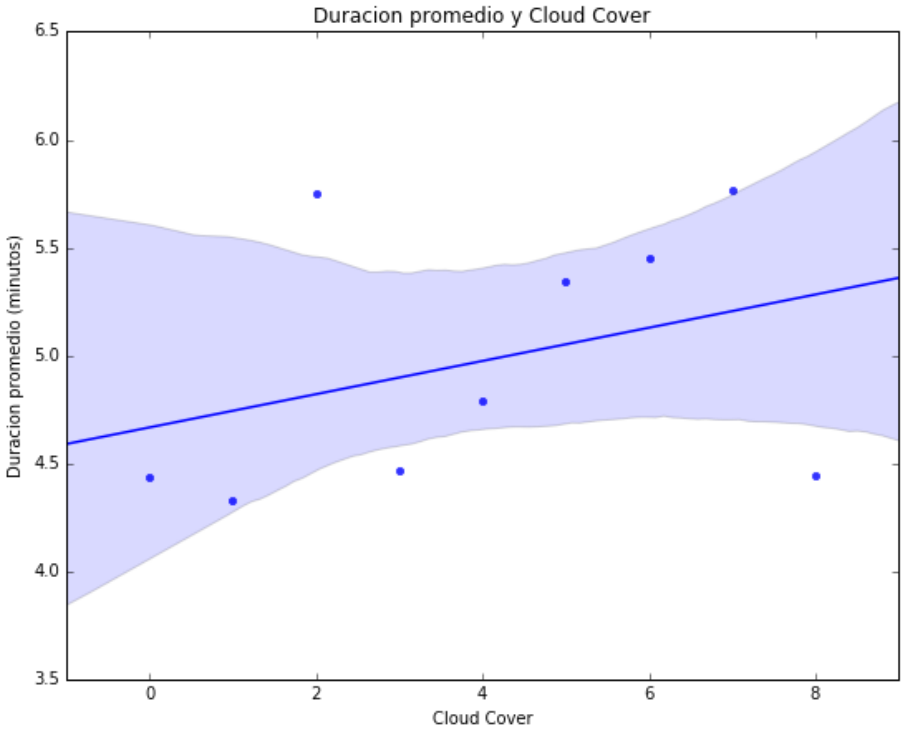



- **¿Hay alguna correlación de precipitation inches y promedio de duración en minutos para el trip más popular?**

  A partir de la siguiente visualización, podemos ver que a medida que aumentan las precipitation inches, dismunuye la duración promedio de los trips. Creemos que esto se puede deber a que los clientes al ver que llueve deciden parar a refugiarse de la lluvia, y por esto se prolonga más el tiempo de viaje.



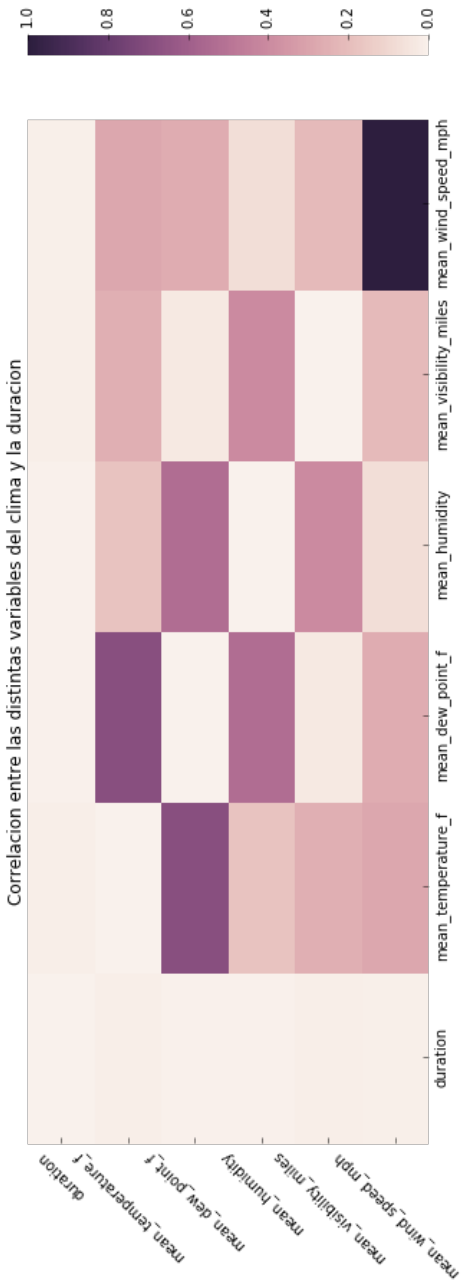

- **¿Hay alguna correlación entre cloud cover y duración promedio en minutos para el trip más popular?**




No pudimos concluir nada a partir de esta visualizacion.

- **Correlación entre las distintas variables del clima y la duración de los trips**

Nuestro objetivo principal fue intentar visualizar cómo se ve afectada la duración de los trips por los distintos factores climáticos, pero vemos que la correlación de la misma con los últimos es muy baja, con lo cual no pudimos extraer ninguna conclusión al respecto.



## Station y Status

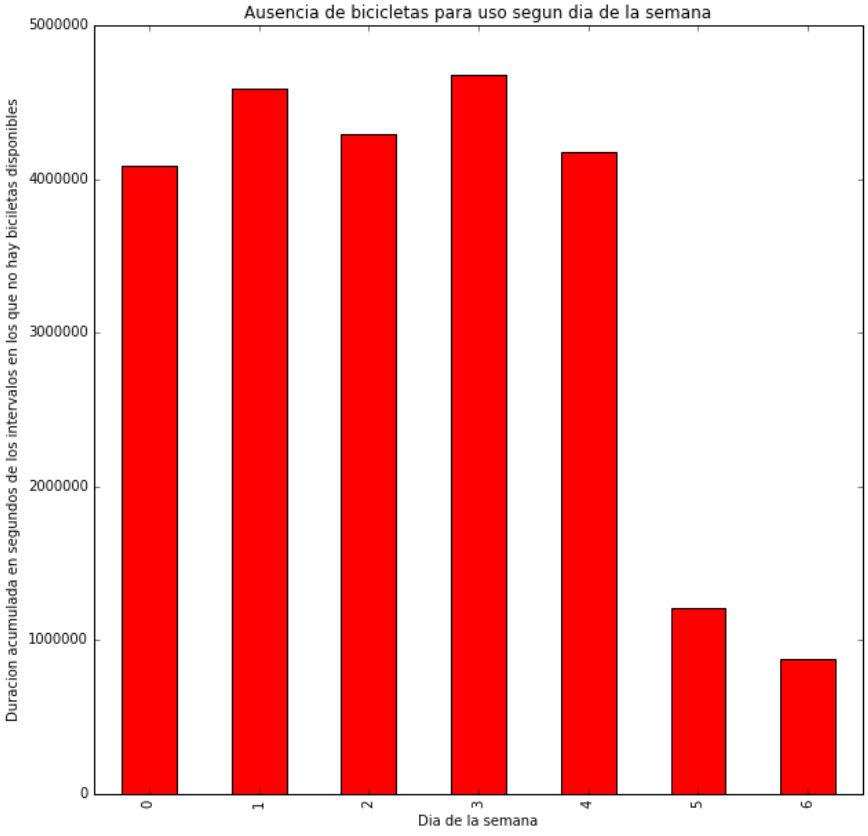
Al iniciar el análisis exploratorio sobre los archivos ‘station.csv’ y ‘status.csv’, nos encontramos con el siguiente inconveniente: el archivo ‘status.csv’ contiene información minuto a minuto del estado de las estaciones, lo cual implica un volumen de almacenamiento de cerca de 2.0 GB. Esto hacía que en algunas de las computadoras de los miembros del equipo fuese imposible levantar el archivo a través de la herramienta Jupyter Notebook, más allá de los múltiples intentos realizados. Por ello, decidimos hacer una leve modificación sobre el data set, con el fin de reducir la cantidad de registros presentes en el mismo (aproximadamente 72 millones). A través de un script en Python recorrimos el archivo CSV original y generamos un nuevo archivo en el que teníamos la información a nivel de intervalos de tiempo, en lugar de minuto a minuto. Se tomaron las condiciones de una estación (cantidad de bicicletas y docks disponibles) en un determinado minuto y se controló hasta qué momento se mantenían las mismas intactas. Entonces se almacenó una línea en el archivo ‘status_red.csv’ con la información para esa estación y los timestamps de inicio y fin de ese intervalo. El script utilizado fue **modificación_status.py**, y se compone de las siguientes instrucciones:

```
#!/usr/bin/env python
import csv

def replace_hyphens (string):
    """Funcion auxiliar para reemplazar guiones por barras en aquellas
    fechas que lo ameriten"""
    if "-" in string:
        return string.replace("-","/")
    return string

def main():
    counter = 0
    with open("status.csv", "r") as f:

        status = csv.reader(f)

        with open("status_red.csv", "w") as d:

            status_red = csv.writer(d)
            status_red.writerow(["station_id", "bikes_available", \
            "docks_available", "start_time", "end_time"])

            station_id_previous = ""
            bikes_available_previous = ""
            docks_available_previous = ""
            start_time_previous = ""
            end_time_previous = ""

            primer_registro = True
            status.next()

            for station_id, bikes_available, docks_available, time in \
            status:

                #counter = counter + 1

                if primer_registro == True:

                    station_id_previous = station_id
                    bikes_available_previous = bikes_available
                    docks_available_previous = docks_available
                    start_time_previous = replace_hyphens(time)
                    end_time_previous = replace_hyphens(time)
                    primer_registro = False

                else:

                    if (station_id_previous == station_id and \
                    bikes_available_previous == bikes_available and \
```

```
                docks_available_previous == docks_available):
                    end_time_previous = replace_hyphens(time)

                else:
                    status_red.writerow([station_id_previous, \
                    bikes_available_previous, \
                    docks_available_previous, start_time_previous, \
                    end_time_previous])
                    station_id_previous = station_id
                    bikes_available_previous = bikes_available
                    docks_available_previous = docks_available
                    start_time_previous = replace_hyphens(time)
                    end_time_previous = replace_hyphens(time)

            #if counter == 70:
                #break

        status_red.writerow([station_id_previous, \
        bikes_available_previous, docks_available_previous, \
        start_time_previous, end_time_previous])

    #print counter


main()
```

De esta manera, el archivo obtenido, 'status_red.csv' consiste en un conjunto de 2 millones de registros, que bien permite trabajar a los datos.

- **Bicicletas agotadas**



Para obtener esta visualización hemos considerado aquellos intervalos de tiempo en los que no había bicicletas disponibles en la estación. Agrupando a los registros por día de la semana, y acumulando la duración de estos intervalos de tiempo obtuvimos el gráfico arriba expuesto.

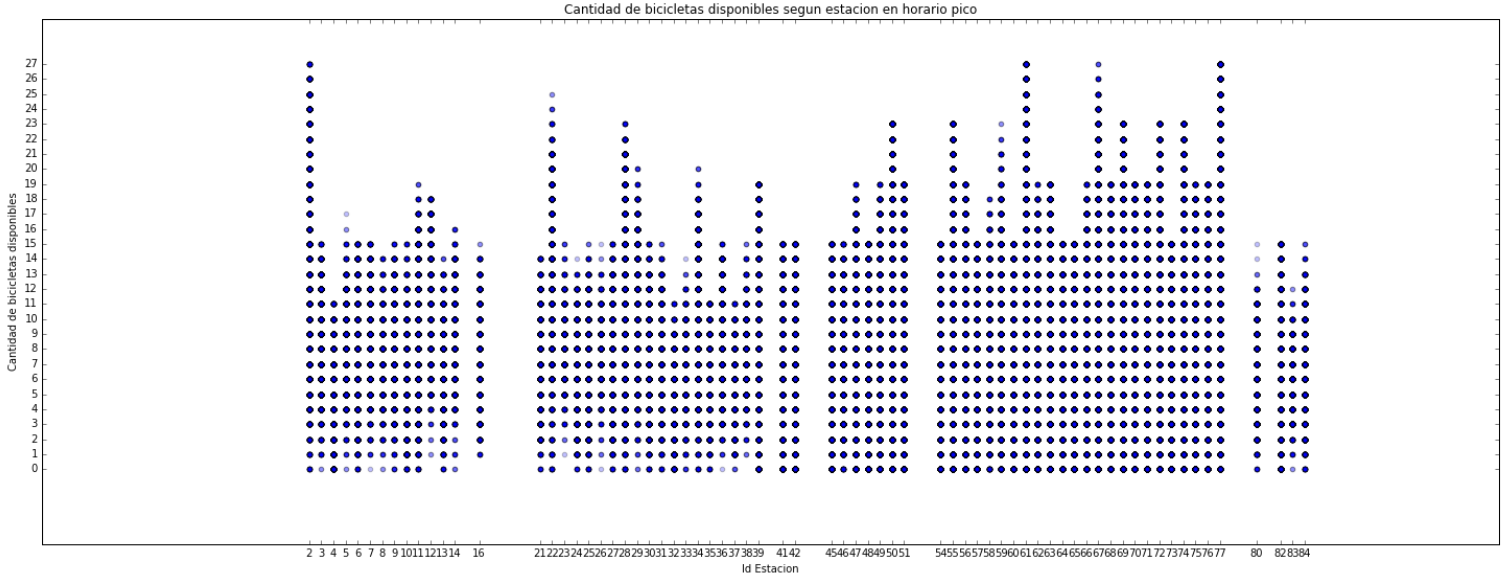
Como se puede ver, durante el fin de semana los valores son considerablemente más bajos dando a entender que las bicis tienden a agotarse mayormente en el rango de días incluido entre lunes y viernes.

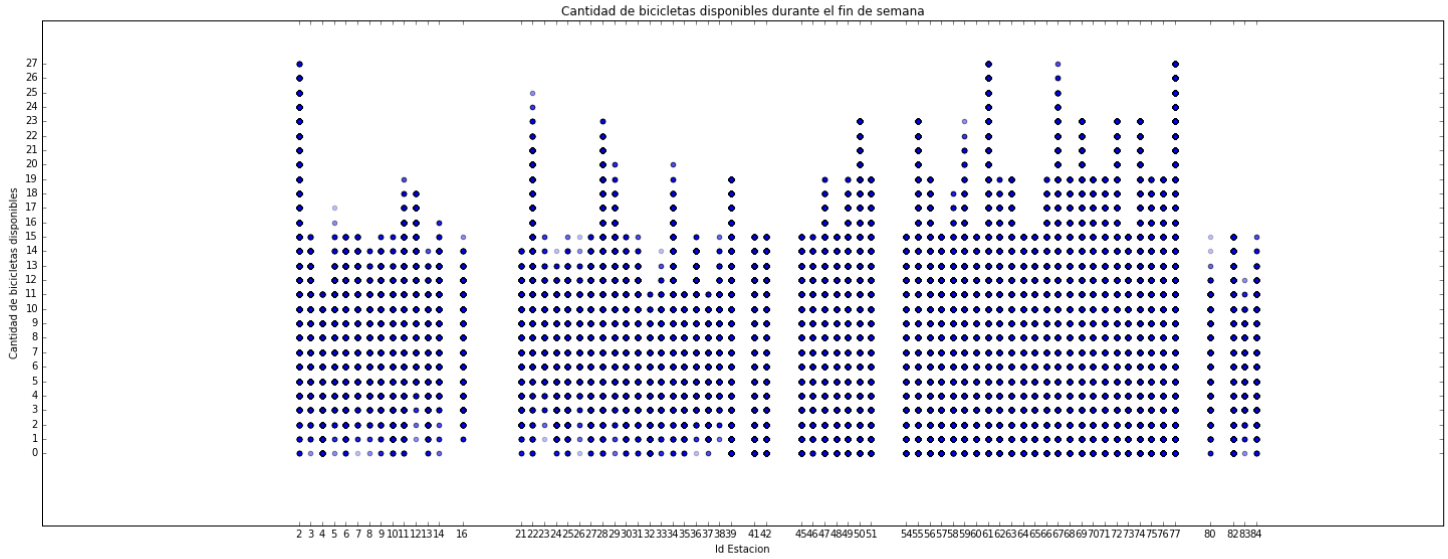
- **Cantidad de bicicletas disponibles durante el horario pico (7:00 - 9:00 y 16:00 – 18:00) y cantidad de bicicletas disponibles durante el fin de semana**

Analizamos qué sucede con la disponibilidad de bicicletas en las estaciones en momentos puntuales de la semana, como ser los intervalos de horario pico (entre las 7 am y las 9 am, y entre las 16 pm y las 18 pm) y los fines de semana.

Comenzando por la fracción de intervalos que están contenidos, total o parcialmente, en el horario pico, destacamos qué hay algunas estaciones en las que es poco frecuente que haya baja disponibilidad de bicicletas, como ser las identificadas por los ids 13, 23, 36, 38. También se observa que es frecuente encontrar bicicletas disponibles en las estaciones, por lo que a primera vista el horario no influiría negativamente en la disponibilidad.

Continuando con el análisis del segundo gráfico, se observan casi las mismas conclusiones para el fin de semana aunque cabe destacar que en muy pocos casos la disponibilidad de bicicletas es cero.


Cantidad de bicicletas disponibles segun estacion en horario pico
Cantidad de bicicletas disponibles
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Id Estacion
2 3 4 5 6 7 8 9 10 11 12 13 14 16 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 41 42 45 46 47 48 49 50 51 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 80 82 83 84


Cantidad de bicicletas disponibles durante el fin de semana
Cantidad de bicicletas disponibles
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Id Estacion
2 3 4 5 6 7 8 9 10 11 12 13 14 16 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 41 42 45 46 47 48 49 50 51 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 80 82 83 84

**Bibliografía y material consultado**

Principalmente:

- Apunte de la cátedra – Capítulo 3: Visualización

- Notebooks de Análisis Exploratorio de la cátedra (link en Piazza)

Algunos links que fueron de ayuda para resolver determinadas dudas:


- http://www.bayareabikeshare.com/open-data (Para ver si podíamos extraer información interesante)

- https://www.unitedstateszipcodes.org/ (Para saber qué zip_code de weather se correspondía con cada ciudad de stations)

- http://seaborn.pydata.org/generated/seaborn.heatmap.html

- http://stackoverflow.com/questions/32723798/how-do-i-add-a-title-to-seaborn-heatmap

- http://seaborn.pydata.org/generated/seaborn.regplot.html

- http://stackoverflow.com/questions/11264521/date-ticks-and-rotation-in-matplotlib