

A report SF Bay Bikers

Group 1

Submitted by:
Sai Srinivas Phaneendra Penumaka
Venkata Bhavana Praveen Pothuri
Artimes Rashidi Torghi
Mohammadreza Rezaei

Content of the Dataset

The Status file contains the data about the number of bikes and docks available for a given station and minute. The Station file contains data that represents a station where users can pick up or return bikes. The Trip file contains data about individual bike trips. The Weather file contains data about the weather on a specific day for certain zip codes.

Weather

Columns

- ◆ Date
- ◆ max_temp_f
- ◆ mean_temp_f
- ◆ min_temp_f
- ◆ max_dew_point_F
- ◆ MEAN_DEW_POI...
- ◆ min_dew_point_f
- ◆ max_humidity
- ◆ mean_humidity
- ◆ min_humidity
- ◆ max_sea_level_pr...
- ◆ mean_sea_level_...
- ◆ min_sea_level_pr...
- ◆ max_visibility_miles
- ◆ mean_visibility_m...
- ◆ min_visibility_miles
- ◆ max_wind_Speed...
- ◆ mean_wind_spee...
- ◆ max_gust_speed...
- ◆ precipitation_inch...
- ◆ cloud_cover
- ◆ events
- ◆ wind_dir_degrees
- ◆ zip_code

Status

Columns

- ◆ Station_ID
- ◆ Bikes_Available
- ◆ Docks_Available
- ◆ Time

Station

Columns

- ◆ station_ID
- ◆ Name
- ◆ Latitude
- ◆ Longitude
- ◆ Dock_Count
- ◆ City
- ◆ Installation_date

Trip

Columns

- ◆ station_id
- ◆ Duration
- ◆ start_date
- ◆ start_station_name
- ◆ start_station_id
- ◆ end_date
- ◆ end_station_name
- ◆ end_station_id
- ◆ bike_id
- ◆ subscription_type

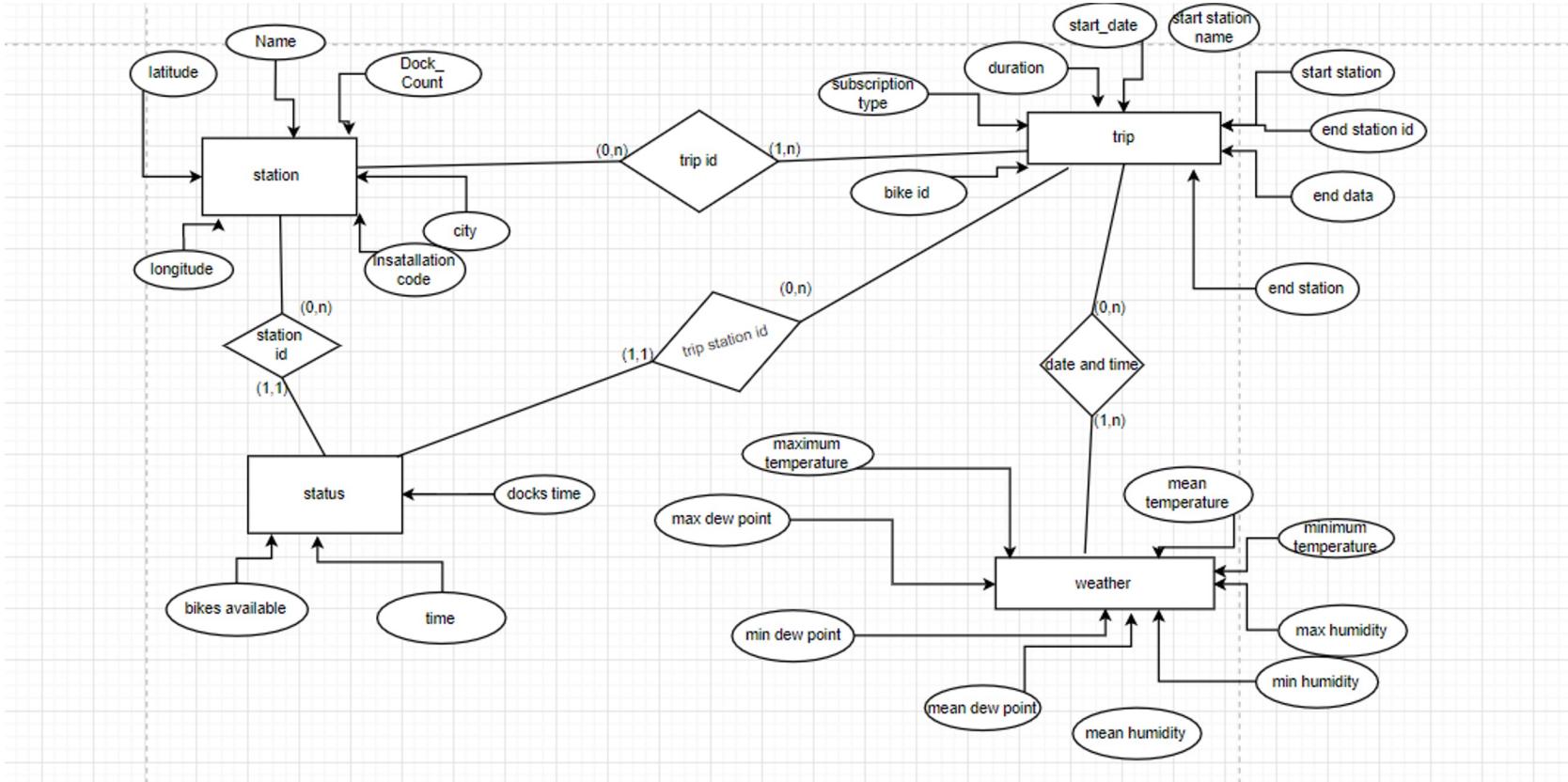
Business Objective

- New business idea
- Analysis of different information
- Ask the correct questions to make sure the business runs smoothly
- Determine the answers for various insights in order to make sure the business has the information it needs.



Bikeshare
The city's bikeshare network is expanding

Conceptual Diagram



Physical Script

```
-- Table `mydb`.`Station`
```

```
CREATE TABLE IF NOT EXISTS `mydb`.`Station` (
  `station_ID` INT NOT NULL,
  `Name` VARCHAR(45) NULL,
  `Latitude` DECIMAL (10,10) NULL,
  `Longitude` DECIMAL (10,10) NULL,
  `Dock_Count` INT NULL,
  `City` VARCHAR(45) NULL,
  `Installation_date` DATETIME NULL,
  PRIMARY KEY (`station_ID`))
ENGINE = InnoDB;
```

```
CREATE TABLE IF NOT EXISTS `weather` (
  `Date` VARCHAR(45) NOT NULL,
  `max_temperature_f` INT NULL DEFAULT NULL,
  `mean_temperature_f` INT NULL DEFAULT NULL,
  `min_temperature_f` INT NULL DEFAULT NULL,
  `max_dew_point_f` INT NULL DEFAULT NULL,
  `mean_dew_point_f` INT NULL DEFAULT NULL,
  `min_dew_point_f` INT NULL DEFAULT NULL,
  `max_humidity` INT NULL DEFAULT NULL,
  `mean_humidity` INT NULL DEFAULT NULL,
  `min_humidity` INT NULL DEFAULT NULL,
  `max_sea_level_pressure_inches` INT NULL DEFAULT NULL,
  `mean_sea_level_pressure_inches` INT NULL DEFAULT NULL,
  `min_sea_level_pressure_inches` INT NULL DEFAULT NULL,
  `max_visibility_miles` INT NULL DEFAULT NULL,
  `mean_visibility_miles` INT NULL DEFAULT NULL,
  `min_visibility_miles` INT NULL DEFAULT NULL,
  `max_wind_Speed_mph` INT NULL DEFAULT NULL,
  `mean_wind_speed_mph` INT NULL DEFAULT NULL,
  `max_gust_speed_mph` INT NULL DEFAULT NULL,
  `precipitation_inches` DECIMAL(10,10) NULL DEFAULT NULL,
  `cloud_cover` INT NULL DEFAULT NULL,
  `events` TEXT NULL DEFAULT NULL,
  `wind_dir_degrees` INT NULL DEFAULT NULL,
  `zip_code` INT NOT NULL,
  PRIMARY KEY (`Date`, `zip_code`),
  ENGINE = InnoDB
  DEFAULT CHARACTER SET = utf8mb4
  COLLATE = utf8mb4_0900_ai_ci;
```

```
-- Table `mydb`.`Trip`
```

```
CREATE TABLE IF NOT EXISTS `mydb`.`Trip` (
  `station_id` INT NOT NULL,
  `Duration` INT NULL,
  `start_date` DATETIME NULL,
  `start_station_name` VARCHAR(45) NULL,
  `start_station_id` VARCHAR(45) NULL,
  `end_date` DATETIME NULL,
  `end_station_name` VARCHAR(45) NULL,
  `end_station_id` INT NULL,
  `bike_id` VARCHAR(45) NULL,
  `subscription_type` VARCHAR(45) NULL,
  PRIMARY KEY (`station_id`),
  ENGINE = InnoDB;
```

```
CREATE TABLE IF NOT EXISTS `Status` (
```

```
  `station_id` INT NOT NULL,
  `bikes_available` INT NULL DEFAULT NULL,
  `docks_available` INT NULL DEFAULT NULL,
  `Date_time` DATETIME NOT NULL,
  PRIMARY KEY (`station_id`, `Date_time`),
  CONSTRAINT `station_id`
    FOREIGN KEY (`station_id`)
    REFERENCES `Bicycle`.`station` (`station_id`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
  ENGINE = InnoDB
  DEFAULT CHARACTER SET = utf8mb4
  COLLATE = utf8mb4_0900_ai_ci;
```

Example of data loading using Terminal

```
chandrikamanikandan@Chandrikas-MacBook-Pro ~ % LOAD DATA LOCAL INFILE '/Users/chandrikamanikandan/Downloads/status.csv'  
INTO TABLE status  
FIELDS TERMINATED BY ','  
ENCLOSED BY '\"'  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS;
```

Insight 1 : How many stations are in each city?

```
2 • select city, count(station_id) as number_of_stations_per_city  
3   from station  
4   group by city  
5   order by number_of_stations_per_city  
6
```

100% 10:1

Result Grid



Filter Rows:

Search

Export:



	city	number_of_stations_per_city
▶	Palo Alto	5
◀	Redwood City	7
◀	Mountain View	7
◀	San Jose	16
◀	San Francisco	35

Insight 2 : Where are most bikes rented from?

```
4 • select name as Station_Name, count(*) as Count from station  
5   inner join trip on station.id = trip.start_station_id  
6   group by Station_Name  
7   order by Count desc limit 10;
```

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows:

Station_Name	Count
San Francisco Caltrain (Townsend at 4th)	49092
San Francisco Caltrain 2 (330 Townsend)	33742
Harry Bridges Plaza (Ferry Building)	32934
Embarcadero at Sansome	27713
Temporary Transbay Terminal (Howard at Beale)	26089
2nd at Townsend	25837
Steuart at Market	24838
Market at Sansome	24172
Townsend at 7th	23724
Market at 10th	20272

San Francisco Caltrain Station (geographical)

Shopping

Union Square

San Francisco station - Muni N-Judah to Powell station

Westfield San Francisco Centre

San Francisco station - 865 Market St. Muni T-Third to Powell station

Downtown Burlingame

Burlingame station - Across from station

Downtown San Mateo

San Mateo station - Two blocks south of station

Hillsdale Shopping Center

Hillsdale station - One block west of station

Downtown San Carlos

San Carlos station - One block west of station

Downtown Redwood City

Redwood City station - One block west of station

Downtown Menlo Park

Menlo Park station - One block west of station

Stanford Shopping Center

Palo Alto station - Quarter-mile west of station

Downtown Palo Alto

Palo Alto station - One block east of station

Downtown Mountain View

Mountain View station - One block west of station

Chase Center



Caltrain is proud to serve the Chase Center, the home of the Golden State Warriors and San Francisco's newest event and concert venue.

Insight 3 : Where are most bikes returned?

```
10 • select name as Station_Name, count(*) AS Count FROM station  
11     inner join trip on station.id = trip.end_station_id  
12     GROUP BY station_name  
13     order by count desc limit 10;  
14
```

< >

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Station_Name	Count
San Francisco Caltrain (Townsend at 4th)	63179
San Francisco Caltrain 2 (330 Townsend)	35117
Harry Bridges Plaza (Ferry Building)	33193
Embarcadero at Sansome	30796
2nd at Townsend	28529
Market at Sansome	28033
Townsend at 7th	26637
Steuart at Market	25025
Temporary Transbay Terminal (Howard at Beale)	23080
Market at 4th	19915

Insight 4 : What is the top 5 maximum duration as hours that a person rented a bike?

```
16 • select round(duration/60/60) AS Duration_Hours  
17   from trip group by duration_hours  
18   order by Duration_Hours desc limit 5;  
19
```

The screenshot shows a database query results grid. At the top, there is a code editor with the SQL query. Below it is a toolbar with various icons for filtering, exporting, and fetching rows. The results grid has a header row labeled "Duration_Hours". The data rows are: 4797, 594, 515, 315, and 201. The second row (594) is highlighted with a light blue background.

	Duration_Hours
▶	4797
	594
	515
	315
	201

Insight 5 : Where are least bikes rented from?

```
21 • select name as Station_Name, count(*) as Count from station  
22     inner join trip on station.id = trip.start_station_id  
23     group by Station_Name  
24     order by Count asc limit 10;  
25 |
```

< >

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows:

Station_Name	Count
Redwood City Public Library	213
Franklin at Maple	224
San Mateo County Center	287
Redwood City Medical Center	311
Mezes Park	341
Stanford in Redwood City	503
Park at Olive	750
Santa Clara County Civic Center	863
California Ave Caltrain Station	1026
Rengstorff Avenue / California Street	1129

Insight 6 : What Bike model is preferred by people who rent bikes?

```
36 • select bike_id, count(*) as Count
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	bike_id	Count
▶	392	2061
	489	1975
	558	1955
▶	267	1951
	631	1948
	518	1942
	532	1933
	592	1932
	395	1927
	368	1926
	540	1926

Insight 7 : Do subscribers rent more bikes, or do customers?

```
25
26      #14. Do 'subscribers' rent more bikes or do 'customers?
27 •   select subscription_type, count(*) AS count
28     from trip
29     group by subscription_type;
30
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

	subscription_type	count
▶	Subscriber	566746
	Customer	103213

Result 5

Read Only

Insights 8 : Which stations do more subscribers rent from vs customers?

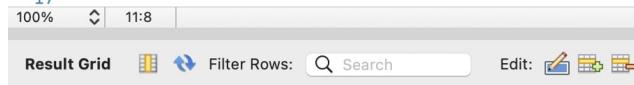
```
5 • select v1.station, Subscriber, SubscriberCount, Customer, CustomerCount, SubscriberCount - CustomerCount difference
6   from
7   (select subscription_type Subscriber, count(*) SubscriberCount, start_station_name station
8     from trip
9    where subscription_type = 'Subscriber'
0   group by subscription_type, start_station_name) V1
1   join
2   (select subscription_type Customer, count(*) CustomerCount, start_station_name station
3     from trip
4    where subscription_type = 'Customer'
5   group by subscription_type, start_station_name) v2
6   on v1.station = v2.station
7   order by SubscriberCount - CustomerCount desc;
```

station	Subscriber	SubscriberCount	Customer	CustomerCount	difference
San Francisco Caltrain (Townsend at 4th)	Subscriber	45842	Customer	3250	42592
San Francisco Caltrain 2 (330 Townsend)	Subscriber	31962	Customer	1780	30182
Temporary Transbay Terminal (Howard at Beale)	Subscriber	25050	Customer	1039	24011
Townsend at 7th	Subscriber	22118	Customer	1606	20512
2nd at Townsend	Subscriber	22482	Customer	3355	19127
Market at Sansome	Subscriber	21213	Customer	2959	18254
Steuart at Market	Subscriber	21465	Customer	3373	18092
2nd at South Park	Subscriber	16945	Customer	1551	15394
Market at 10th	Subscriber	17548	Customer	2724	14824
Harry Bridges Plaza (Ferry Building)	Subscriber	23502	Customer	9432	14070

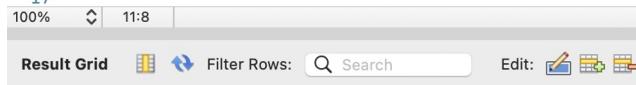
station	Subscriber	SubscriberCount	Customer	CustomerCount	difference
University and Emerson	Subscriber	385	Customer	1127	-742
San Jose Civic Center	Subscriber	764	Customer	830	-66
San Jose Government Center	Subscriber	12	Customer	11	1
California Ave Caltrain Station	Subscriber	521	Customer	505	16
Broadway at Main	Subscriber	46	Customer	21	25
Franklin at Maple	Subscriber	145	Customer	79	66
Redwood City Public Library	Subscriber	142	Customer	71	71
Park at Olive	Subscriber	414	Customer	336	78
San Mateo County Center	Subscriber	237	Customer	50	187
Mezes Park	Subscriber	290	Customer	51	239

Insight 9 : How many bikes are available at each station?

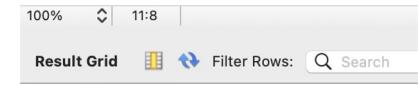
```
select station_id, bikes_available  
from status  
15 group by station_id  
16 order by station_id  
17
```



```
select station_id, bikes_available  
from status  
15 group by station_id  
16 order by station_id  
17
```



```
13 • select station_id, bikes_available  
from status  
14 group by station_id  
15 order by station_id  
16  
17
```



Insight 10 : On which days of the week do customers and subscribers rent the most bikes?

```
7 • select count(*) as trips_count, dayname(start_date), subscription_type from trip  
8 group by dayofweek(start_date),subscription_type order by dayofweek(start_date);  
9  
10  
11
```

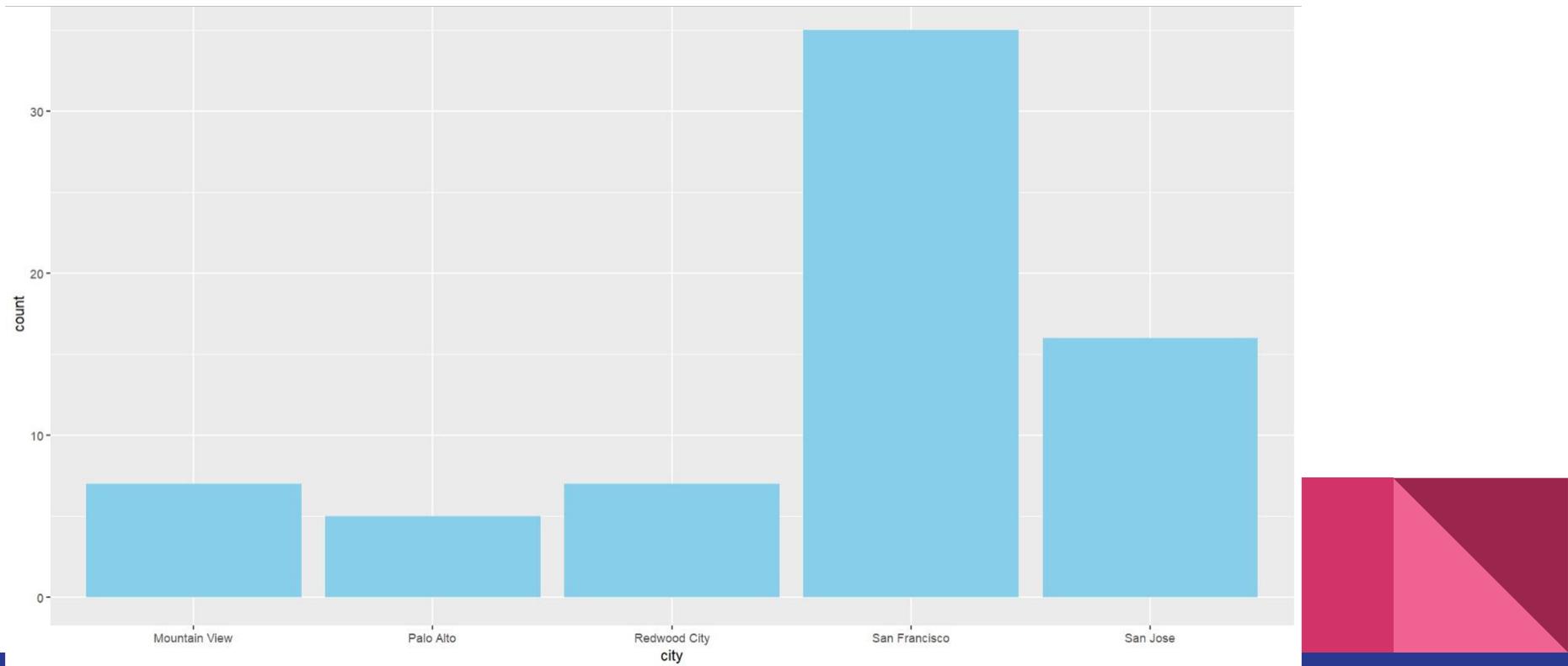
Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

trips_count	dayname(start_date)	subscription_type
19687	Sunday	Customer
18704	Sunday	Subscriber
11469	Monday	Customer
104404	Monday	Subscriber
11040	Tuesday	Customer
111219	Tuesday	Subscriber
11495	Wednesday	Customer
108706	Wednesday	Subscriber
12451	Thursday	Customer
106638	Thursday	Subscriber
14946	Friday	Customer
94415	Friday	Subscriber
22125	Saturday	Customer

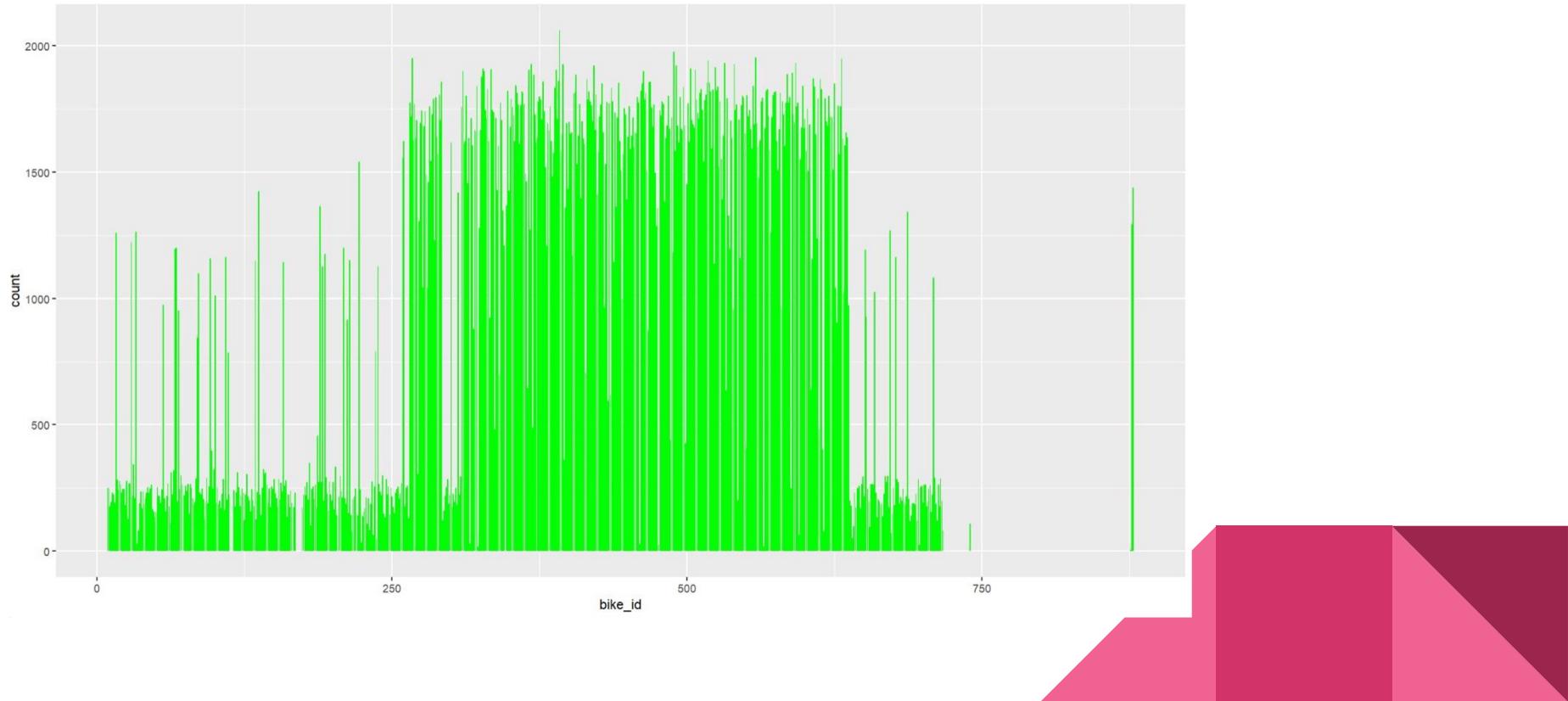
Result 2 ×

Result Grid | Form Editor | Field Types | Query Data | Read Only

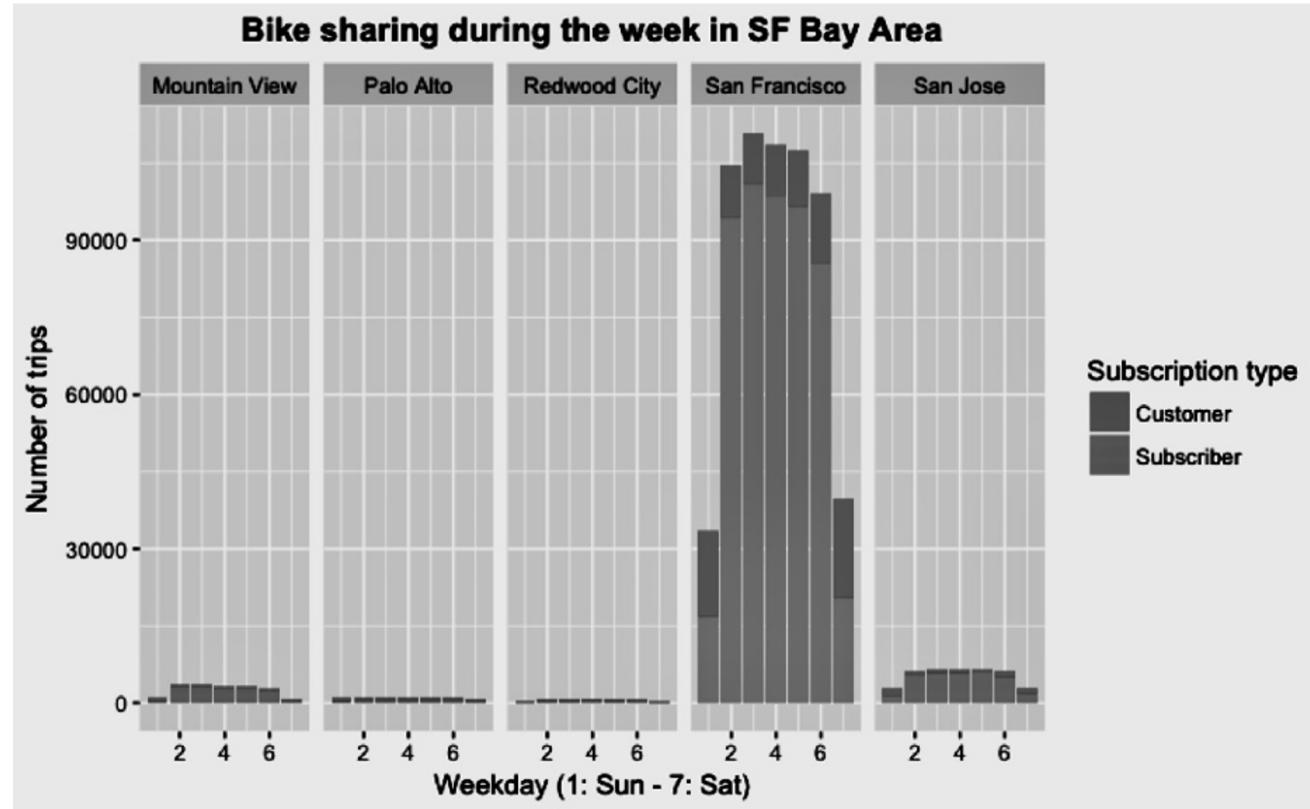
Visualization Charts - How many stations are in each city?



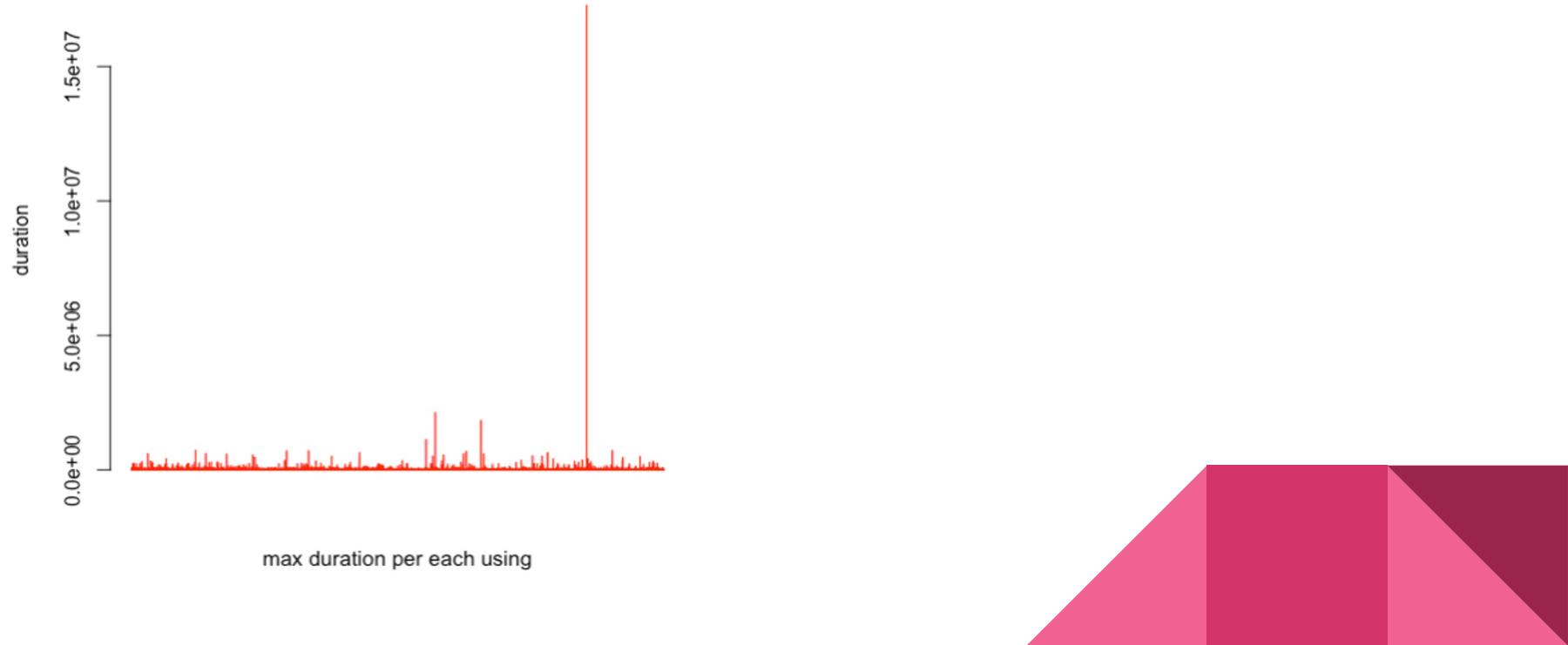
Visualization Charts- What model of bike is preferred by customers?



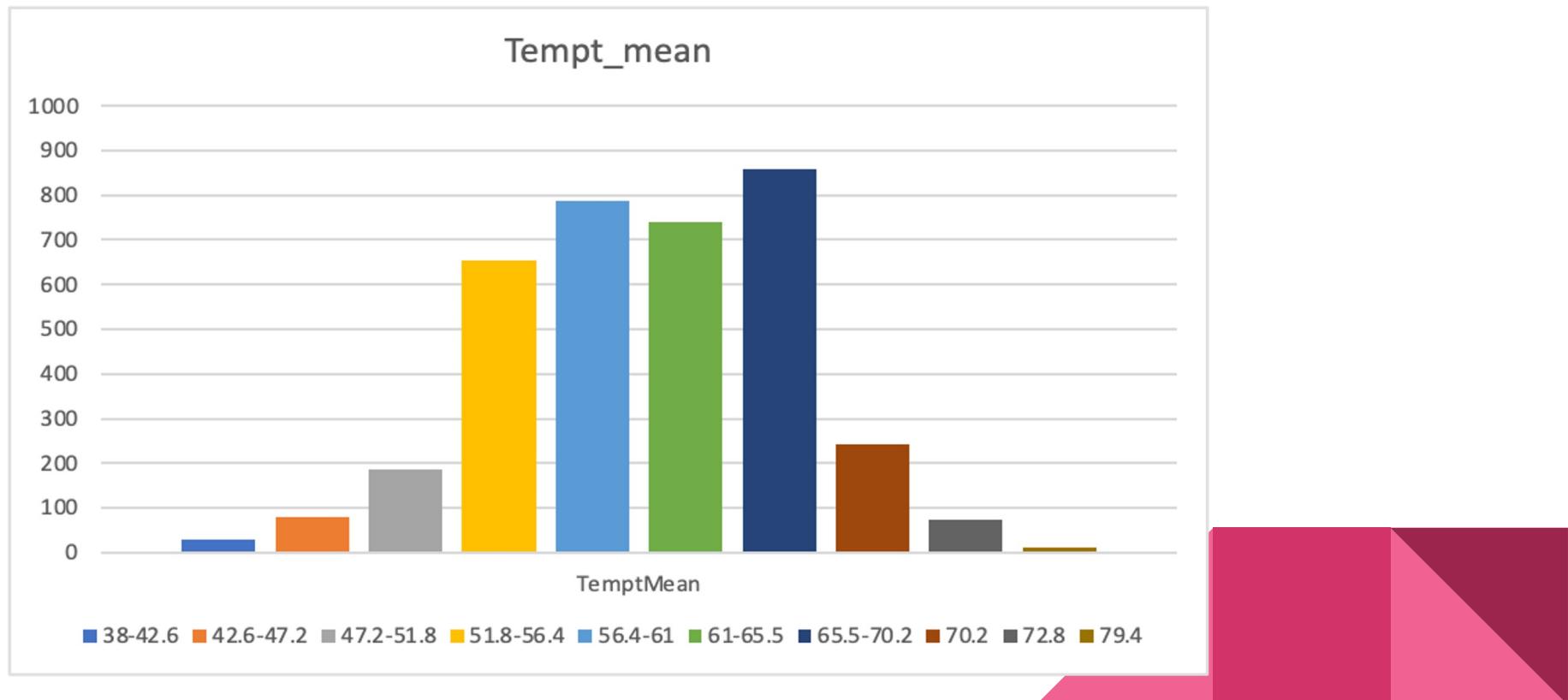
Visualization Charts- What days of the week are most bikes rented on?



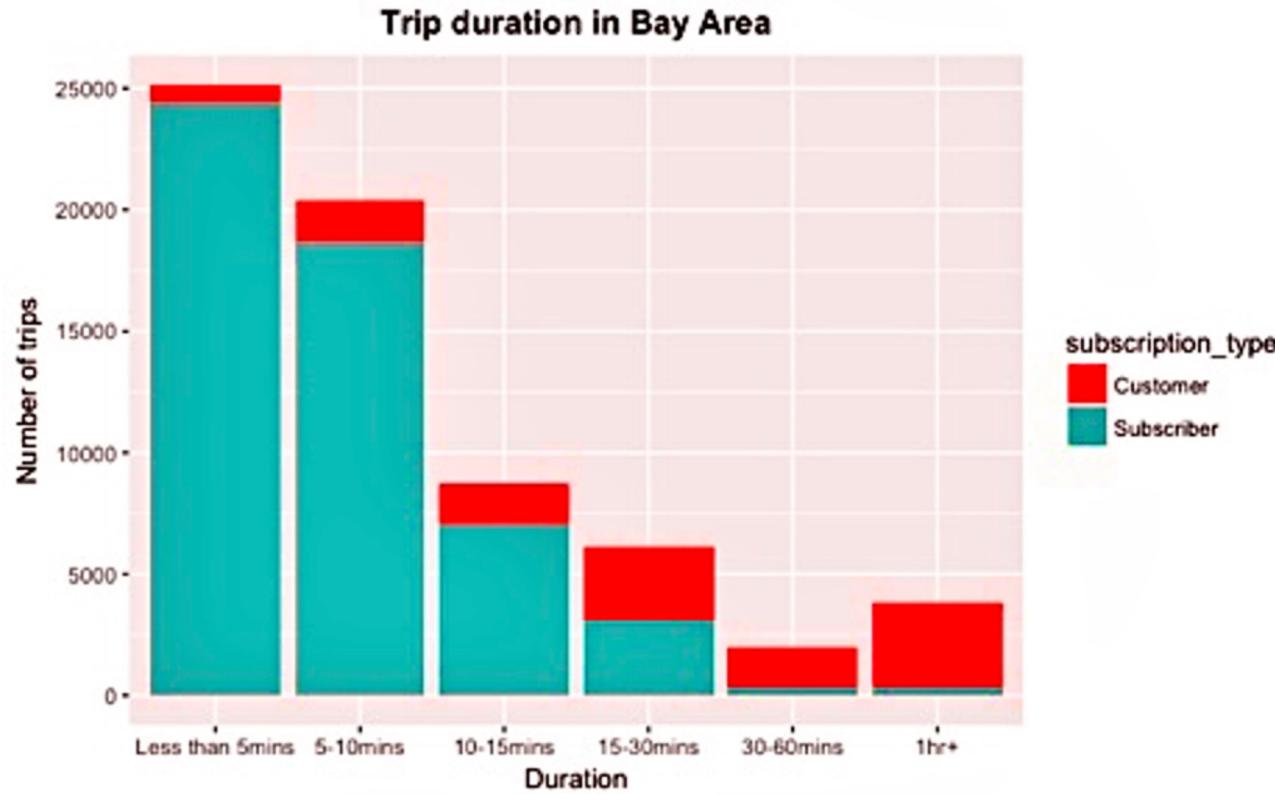
Visualization Charts- How long do customers rent bikes?



Visualization Charts- What is the average temperature on days with the least and most rented bikes?



Visualization Charts- How long do customers rent bikes vs subscribers?



Recommendations

- SF Bay Bikers is a lucrative business
- Focus more on subscribers as they rent more
- Position more stations around popular tourist attractions and train stations
- Provide different types of bikes according to terrain
- Weekend incentives
- Expand into different skill groups (tricycles, unicycles)

Thank you

Any Questions?