Bike Sharing Analysis

Group 7:

Honghui Hu

Nagarjuna Lokaraju

Fatima Nurmakhamadova

Yu Qiu

Xiaolu Shen

Prof. Jack Bergersen ALY 6070 - 21800

Communication and Visualization for Data Analytics

College of Professional Study, Northeastern University

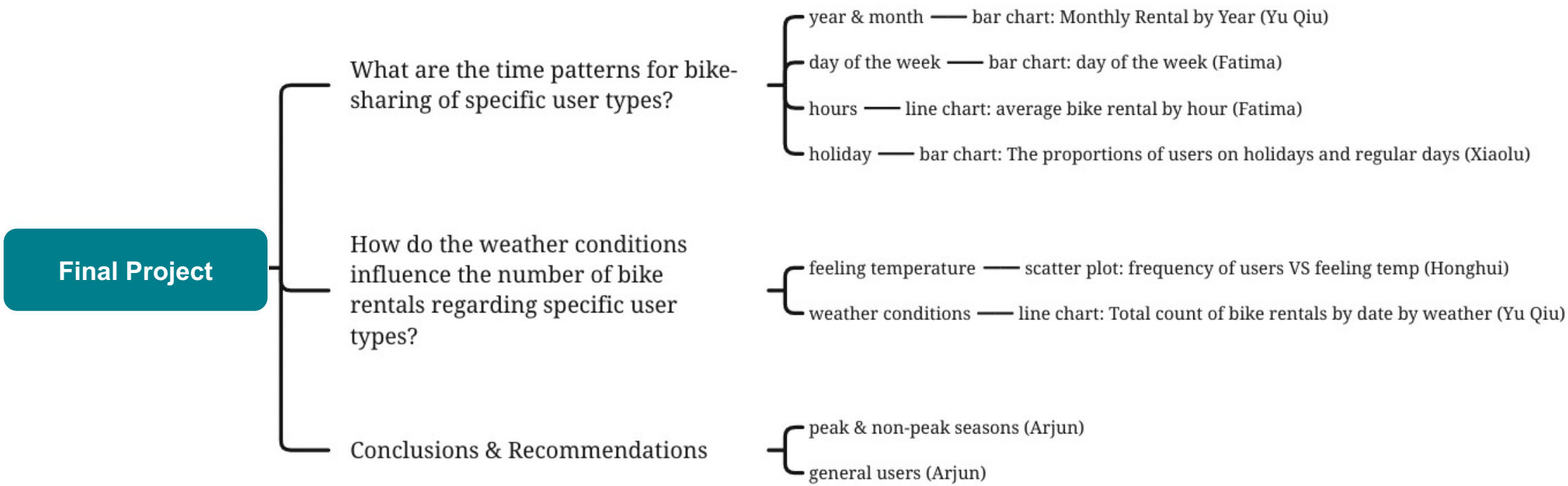
March 29, 2022

Introduction

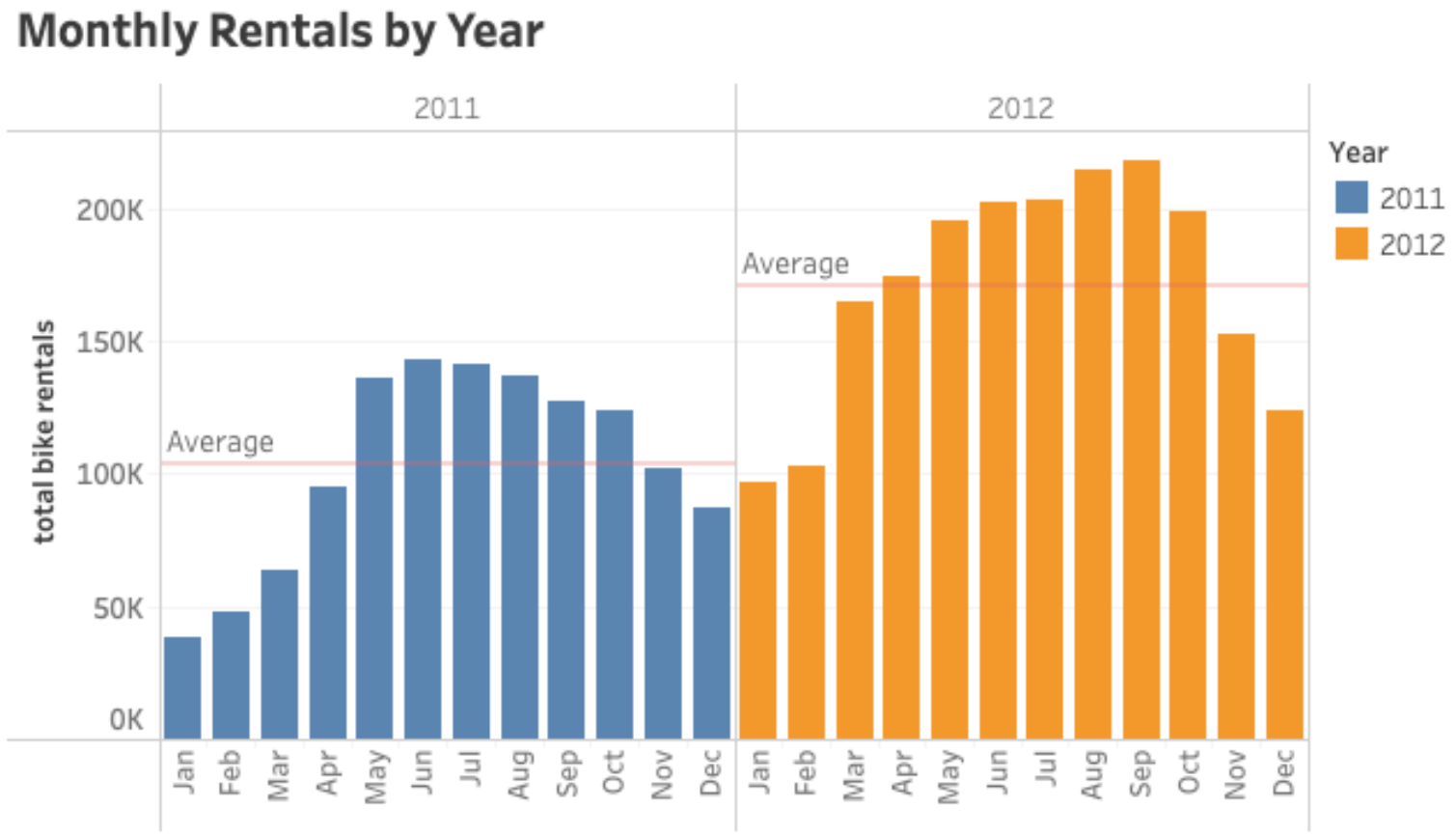
Purpose

The **purpose** of this project is to provide information to product managers based on the analysis of

- demand for bike rentals by different factors like months, holidays, hours, weather condition, temperature, etc.
- difference in the demand of registered users and casual users



Users Frequency Distribution Analysis



Findings

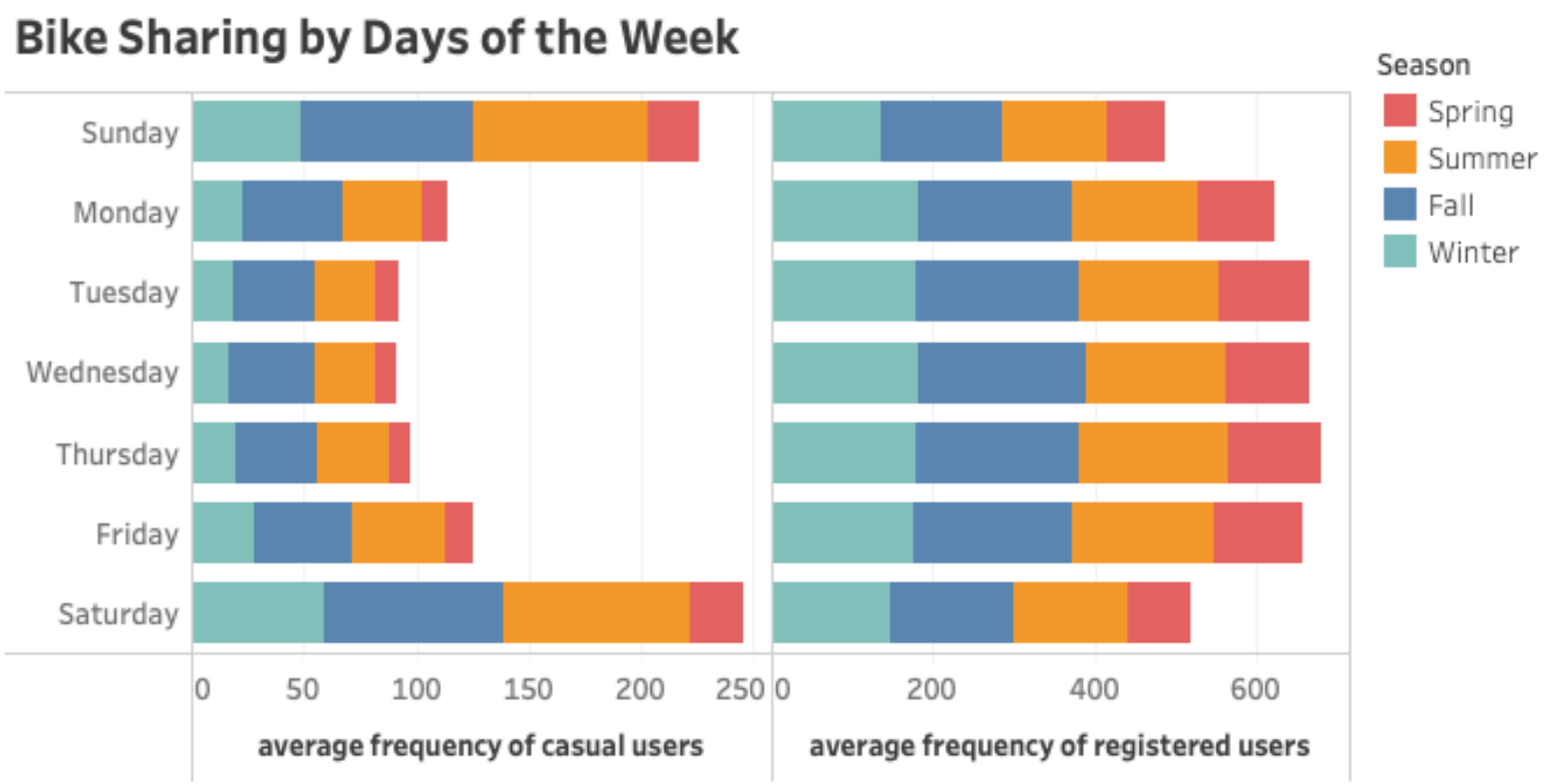
- Rentals **increased** in 2012
  - Average line higher in 2012
  - Month to month growth
  - One more month above average line in 2012
- Rentals from **May to September higher** than annual average in both years
- **Peak Month**
  - Peak month in 2011: June
  - Peak month in 2012: September
- **Low Demand Period**: Jan and Feb

Reasons to use bar chart with reference lines

Bar charts could quickly show the difference in values for related data sets side by side.

- In this case, we compared the total rentals in each month and by using a bar chart, we could also effectively observe the distribution pattern in a year.
- Added average **reference lines** to show in which months the total rentals are very high and compare the 2 reference lines in 2 years to compare the general trend.

Users Frequency Distribution Analysis



Findings

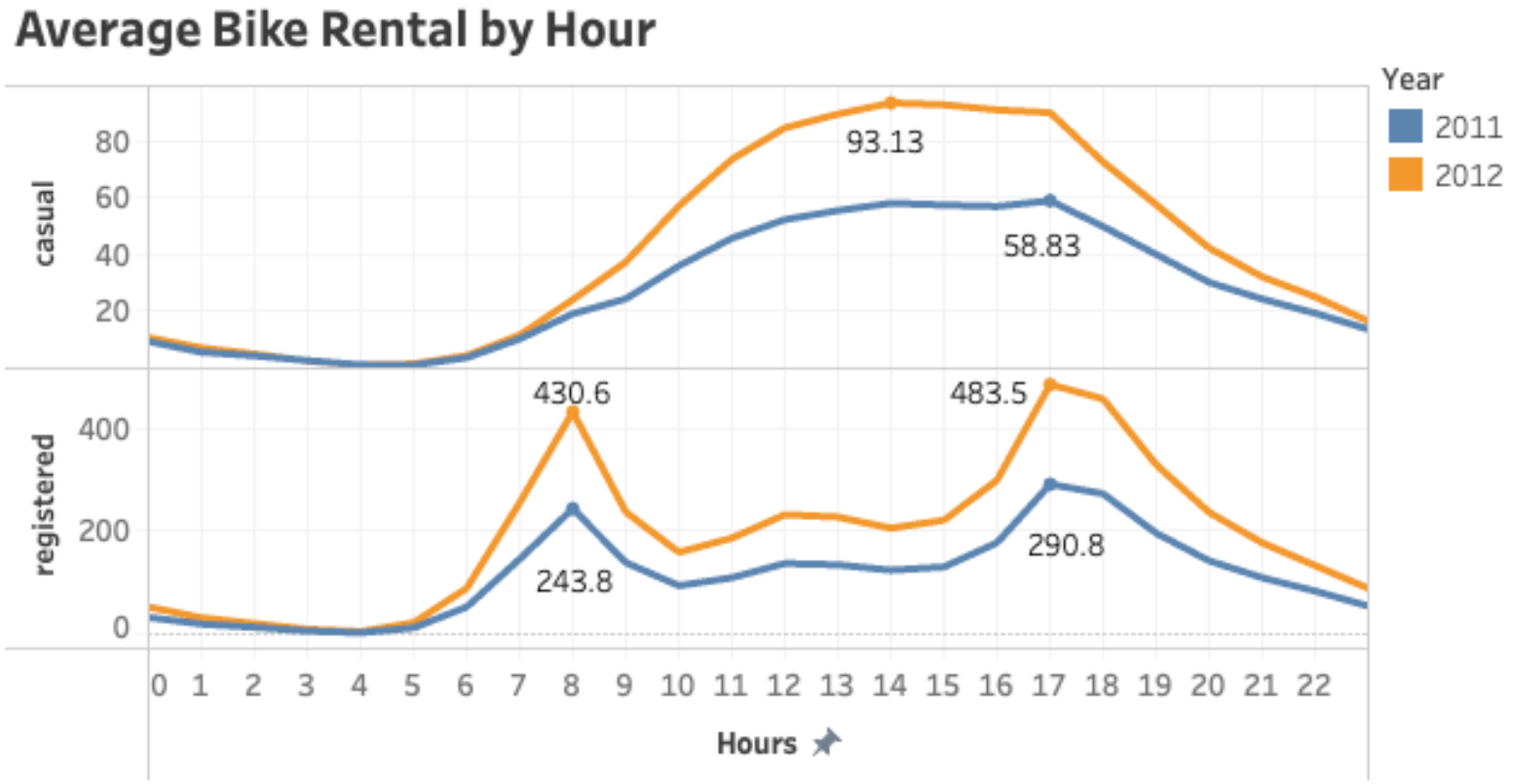
- Casual users tend to rent mostly on weekends. The peak day is Saturday - 250 users
- Registered users tend to rent mostly on weekdays. The peak day is Wednesday - 690 users
- Spring is less demanded season by both users

Reasons to use bar chart

Bar charts helps to compare each user type of bike-sharing on a different day of the week which is categorical data.

- In this case, we compared the average rentals in each day of the week, thus can see the demand trend for each user type.
- Coloring the bars by season helps to compare the difference in users' size in different seasons

Users Frequency Distribution Analysis



Findings

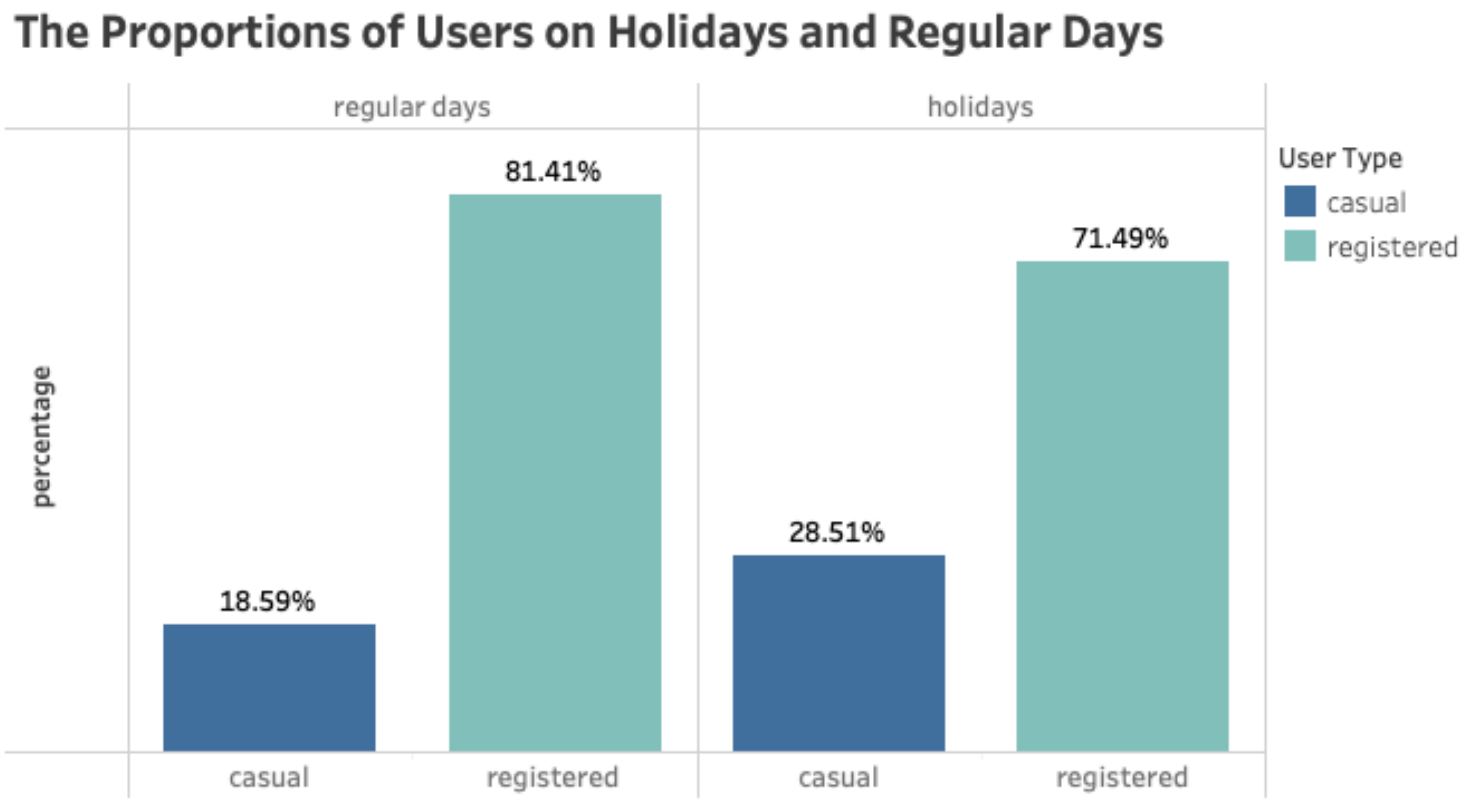
- Almost the same behavioral pattern of bike-sharing by hours in different years (both users)
- Casual users tend to rent bikes mostly during the afternoon
- Registered users tend to rent bikes mostly in the morning and evening
- The time-period from midnight to early morning and late afternoon is unpopular

Reasons to use line chart

Line charts can demonstrate possible changes and trends over the period, allowing to compare the users' distribution within both years

- In this case, we compared the average number of users by hour, and labeled the peaks of the distribution.
- Comparing the sum of users might increase the cognitive load of the viewers as numbers are higher to quickly perceive

Users Frequency Distribution Analysis



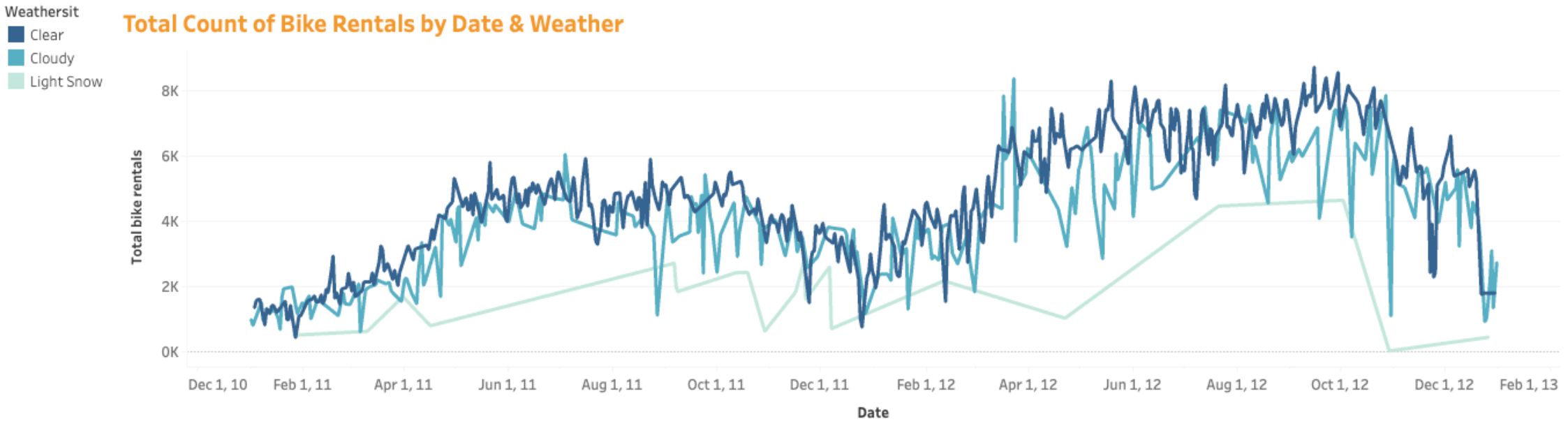
Findings

- Casual users on holidays:22,359;
  on regular day: 597,658.
- Registered users on holidays:56,076;
  on regular days:2,616,586.
- **The frequency of casual users shows an obvious increase on proportion during holidays**

Reasons to use bar chart

The reason we chose this visualization form is that the bar chart can clearly show the possible differences existing in the different groups and dimensions.

Time Series Analysis of Total Bike Rentals



Findings

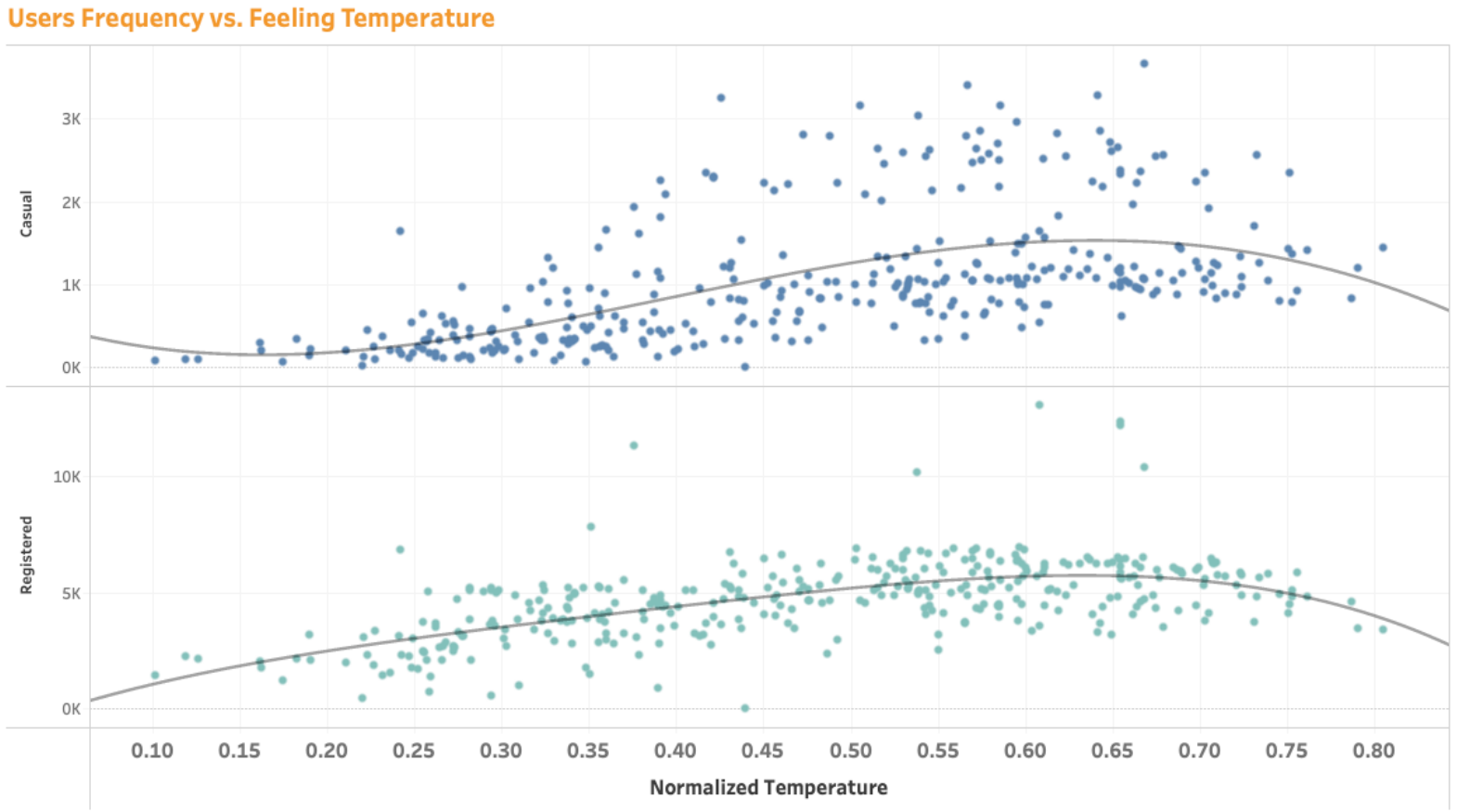
- Seasonality of **clear day** and **cloudy day** are similar
  - a **peak** from **June to September**
  - a **valley** from **December to February**
- **Hard to predict on light snowy days:** an obvious seasonality or trend
- Light snowy days had fewer bike rentals

Reasons to use line chart

Line charts could effectively show the trend over time.

- The above analysis is to see the changes of total bike rentals over each day from 2011 to 2012.
- 3 lines to tell the difference in trend with different weathers

Users Frequency with Different Feeling Temperature



Findings

- For registered users, there is distinct correlation between rental frequency and temperature.
- For casual users, there is no obvious pattern to see the correlation between rental frequency and temperature.
- A high frequency casual rentals (widespread) between 0.30 and 0.70 temperature interval.

Reasons for scatter plot with polynomial model

Scatter plot allows the audience to catch the user frequency distribution on normalized feeling temperature.

Polynomial model provides a better approximation of the relationship between the feeling temperature and the number of users than simple linear regression.

Recommendations

Peak Seasons

Increase availability of bikes to maximize profit:
- good weather conditions: between 0.25 to 0.65
- peak hours of 7am and 5pm
- peak period: May to October

Non- Peak Season

Spring:
- Conduct marketing, decreasing prices to attract and improve the bike sharing
- Reduce the availability to decrease the cost

General Usage

Offer extra service for casual users during the holidays, and do a 3-day trial etc. to bring new customers, and maintain the existing
Do predictions to help the business grow and be accurate on the profit and cost effectiveness

Thank you!