**Final Project: Forecasting Bikeshare Usage for the City of Austin**

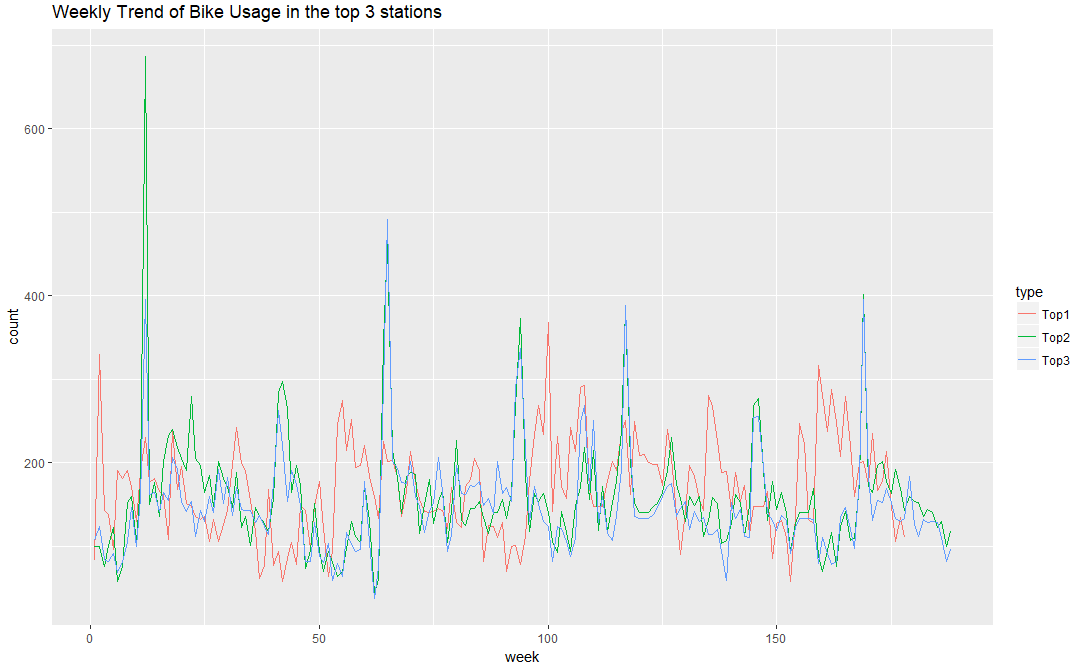
Oliver Yuan, Ruyi Shi

**Business Problem:**

As bike shares are becoming a popular alternative means of transportation in many cities. The City of Austin is trying to identify top 3 and bottom 3 stations as well as forecast future bike usage so that they meet demand at stations.

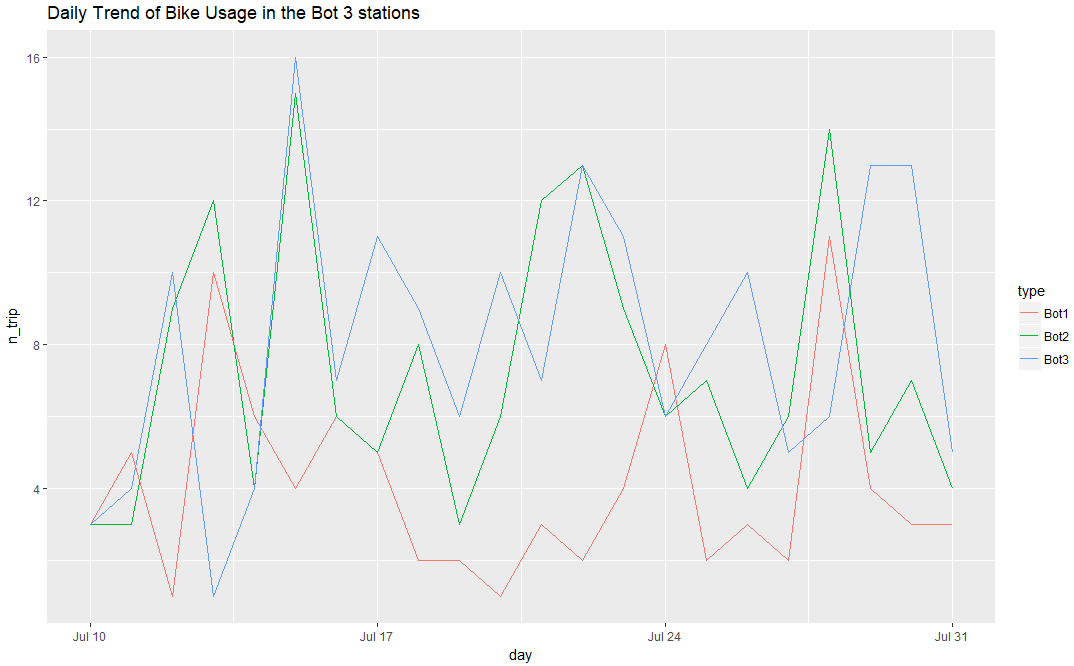
**Analysis:**

After aggregating the data by station name and calculating the number of trips initiated in each station. We found that the top 3 stations are: **Riverside @ S. Lamar (28695 trips), City Hall / Lavaca & 2nd (28535), 5th & Bowie (26669).** Below is a chart for these 3 stations:



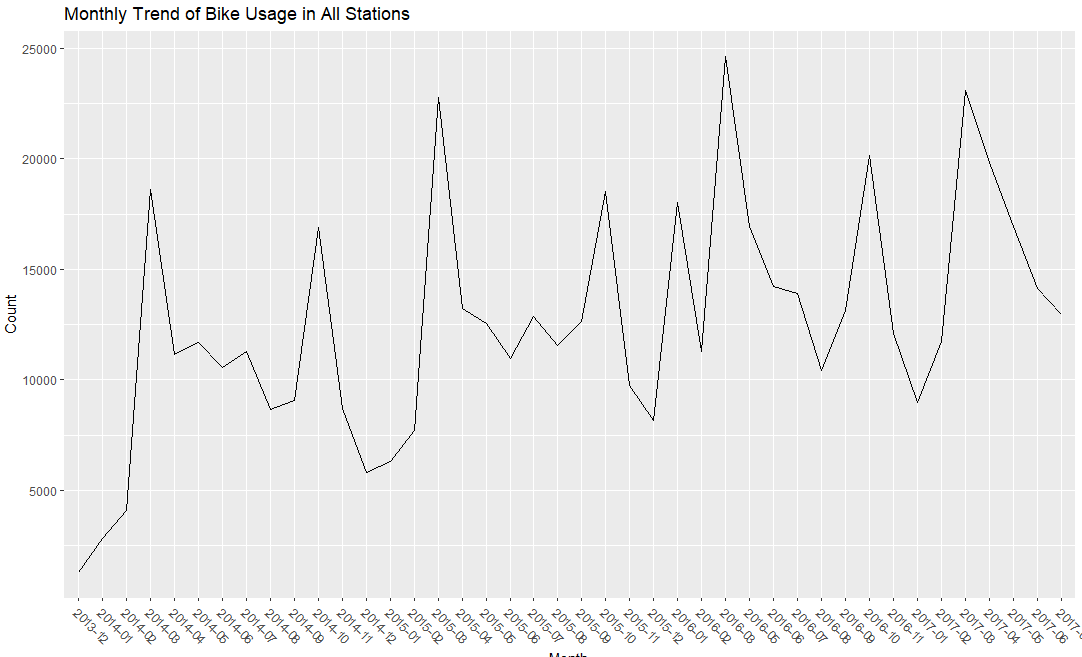
As we can see in the plot, the bike usage for the 3 stations is pretty stable along with the time expect several peak weeks. Furthermore, **station Riverside @ S. Lamar and station 5th & Bowie have the same peak weeks and their trends look very similar**. Station City Hall / Lavaca & 2nd is different from other stations in terms of overall trend and peak time. **Therefore, it is important to figure out what happened during those peak weeks and plan bike supply ahead.**

We also found that the bottom 3 stations are: **Henderson & 9th (96 trips), Congress & Cesar Chavez (182 trips) and Sterzing at Barton Springs (198 trips).** Below is a chart for the 3 stations:

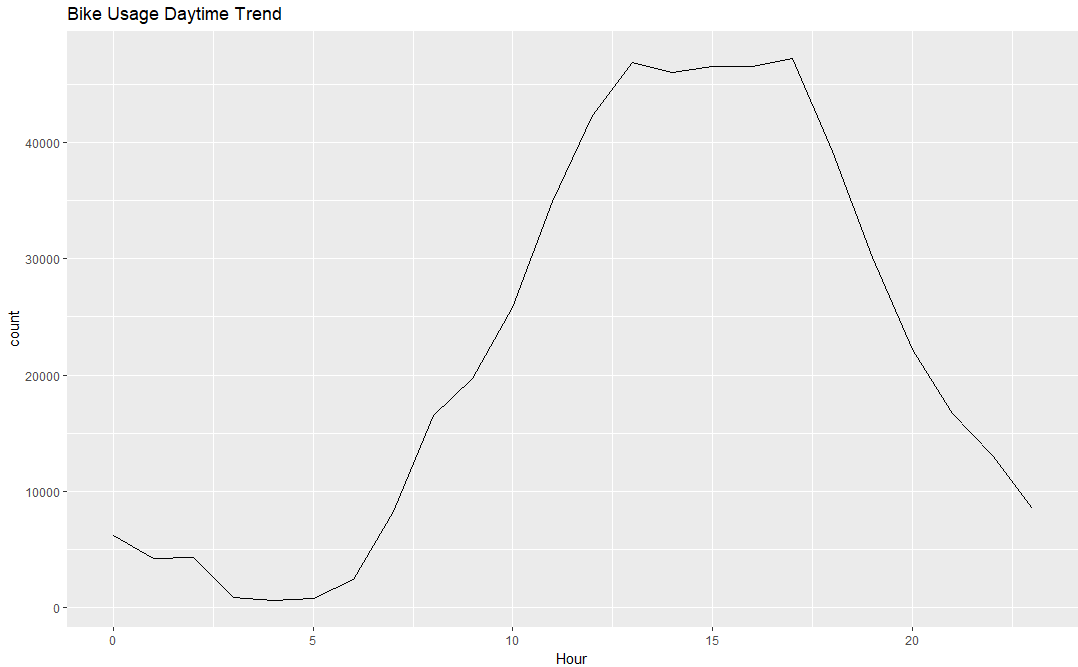


The reason that these stations are the bottom 3 is that they are open for only a couple of weeks. And the range of bike usage is from 0 to 16, which is a relative small range. Therefore, I will suggest that we investigate these stations longer before making any decisions.

Then, we plotted the overall trend of bike usage in all stations:



It is interesting that bike usage will soar in **March** and **October** in each year. And we plotted daytime bike usage.



We see that bike usage centralized from **13:00** to **17:00**.

To sum up, we will recommend that the City of Austin figure out what happened during peak daytime, week and month so that they can plan bike supply better. Furthermore, they need to give more time to the bottom 3 stations since they are open recently.

**Forecast for top 3 stations:**

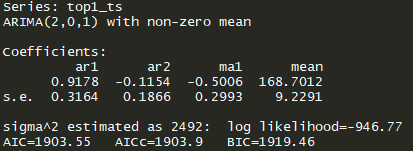
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Riverside @ S. Lamar | 139 | 148 | 153 | 157 | 160 | 162 |
| City Hall / Lavaca & 2nd | 149 | 106 | 107 | 128 | 156 | 154 |
| 5th & Bowie | 101 | 70 | 56 | 102 | 117 | 107 |

**Appendix:**

Before building models, we have to aggregate data into weekly level. We first arranged data by date and figured out the range of date for each series. Then we create vectors that contain complete date range and used data right join with that table. Then we imputed missing values by taking the median of the series.

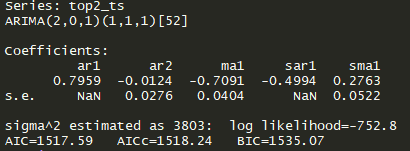
After cleaning the data, we plotted data, did Ljung-Box test to check if the series is white noise and did Dickey-Fuller test to see if the series is stationary. Then we tried ARIMA, simple exponential smoothing and HoltWinter’s method. It ended up that ARIMA is the best model for all 3 series.

Model for top 1 station Riverside @ S. Lamar:





Model for top 2 station City Hall / Lavaca & 2nd:





Model for top 3 station 5th & Bowie:

