

Explanatory Data Analysis on Washington D.C. Bike Data

PRESENTATION BY: TANNER PITTS

Why is this Data Important?

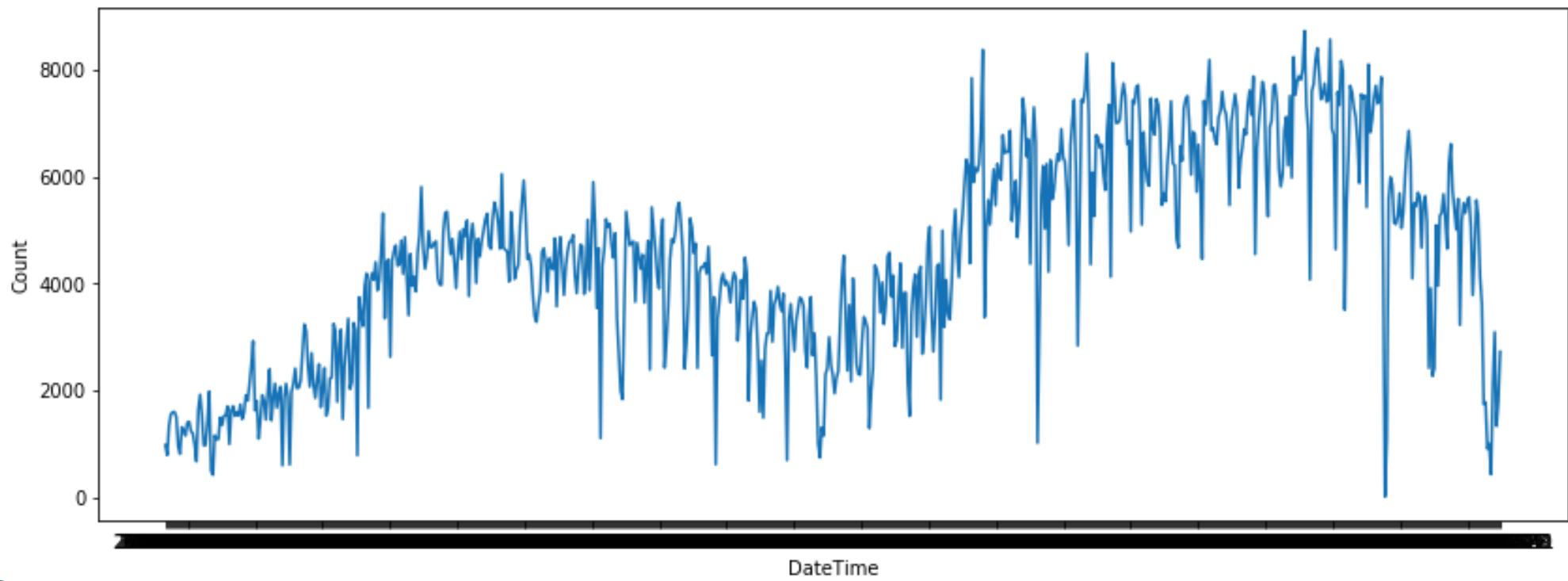
- ▶ We need to make sure that the city has the proper number of bikes to deal with demand
- ▶ Make sure that during slow months proper number of bikes are stored to reduce wear and tear.



The Data

731 Observations

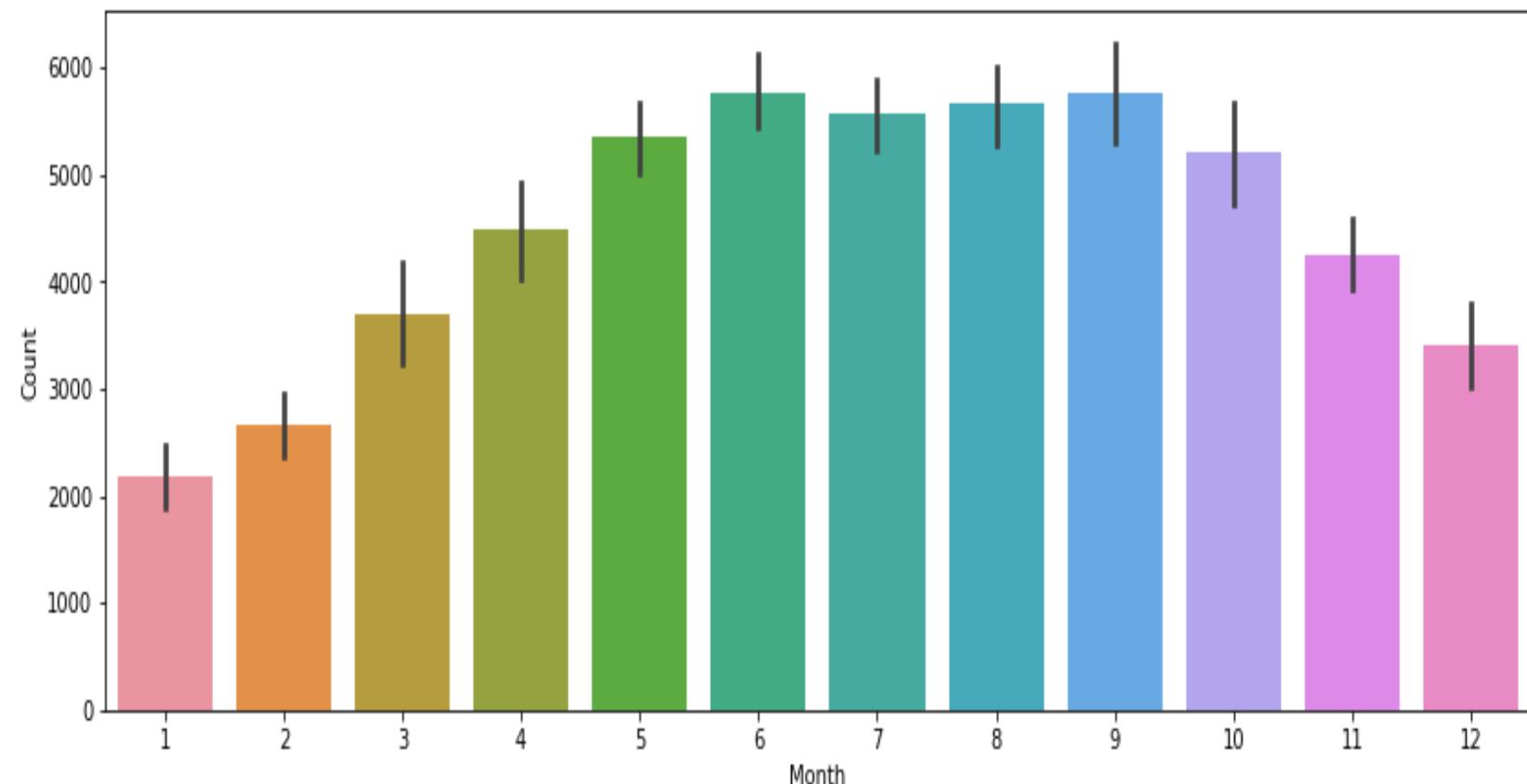
Variable	Units	Definition
DateTime	yyyy/mm/dd	Date of the given data
Count	Bikes	Count of total bikes used
Holiday	Boolean	1: Holiday 0: Not
Temp	Celcius/41	The average temperature
Hum	% out of 100	The average humidity
DayofWeek	Categorical	Day of the week
2-12	Boolean	1:Given Month 0:Not
Sat-Thu	Boolean	1:Given Day 0:Not
2012	Boolean	0: 2011, 1:2012
Temp2	Farenheit	The average temperature

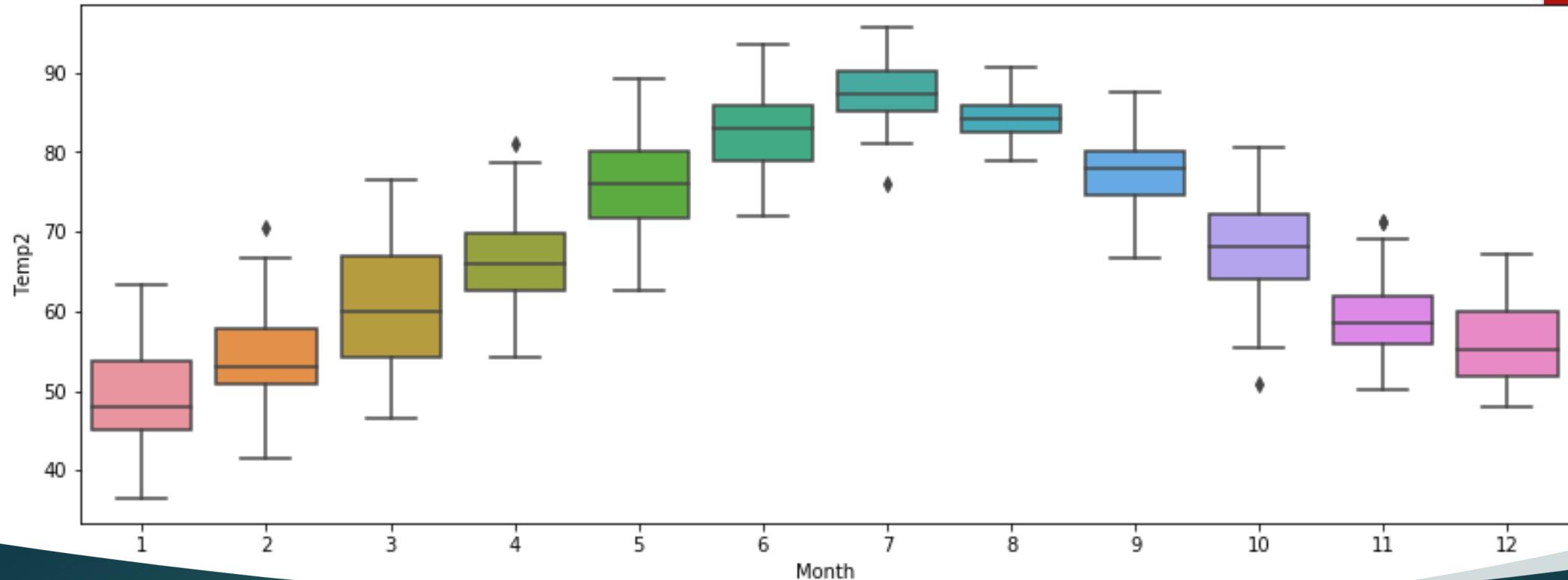


Change in Count over Time

Change in count by Month

- Highest usage months are between May and October

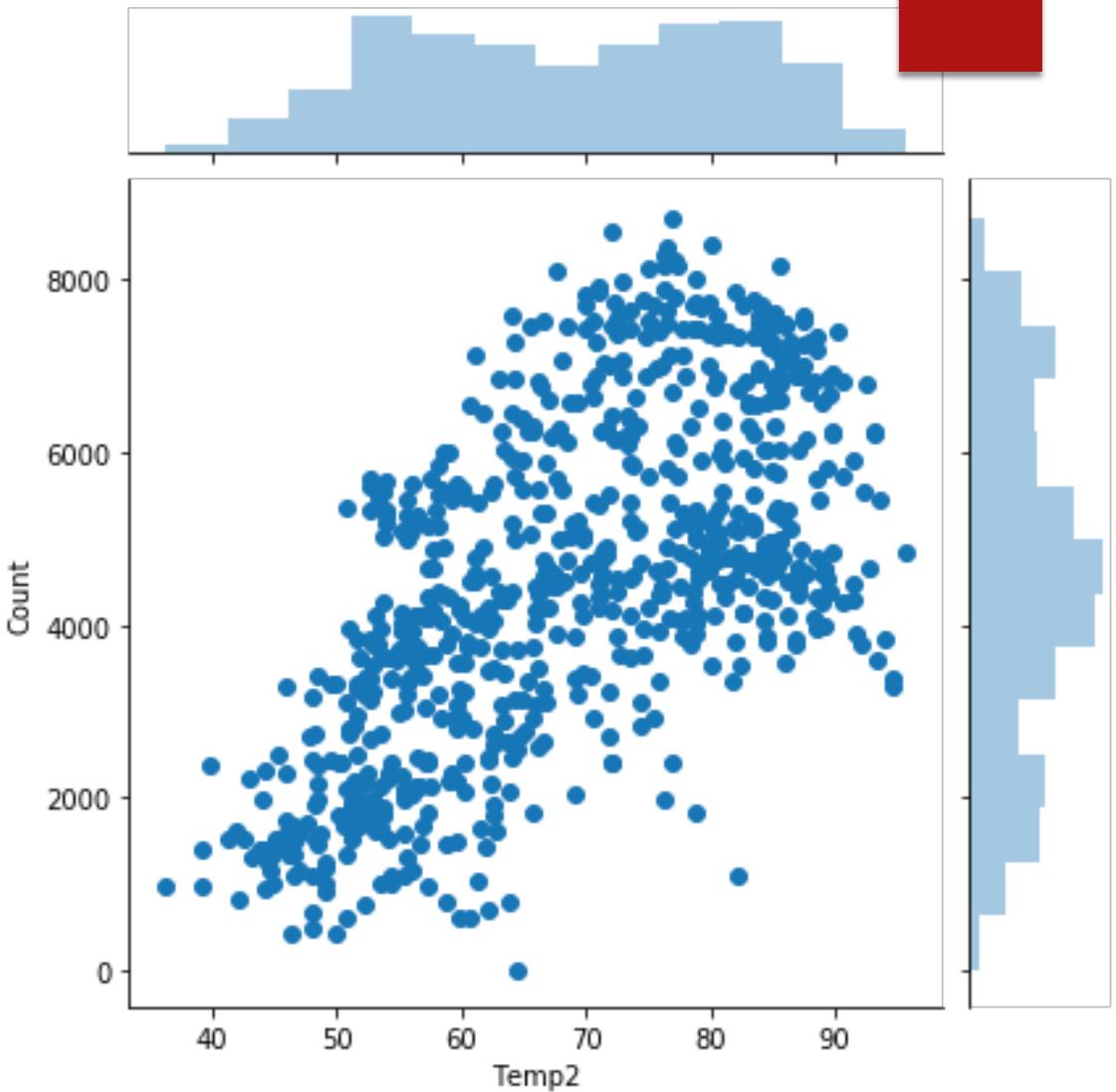




Change in Temperature by Month

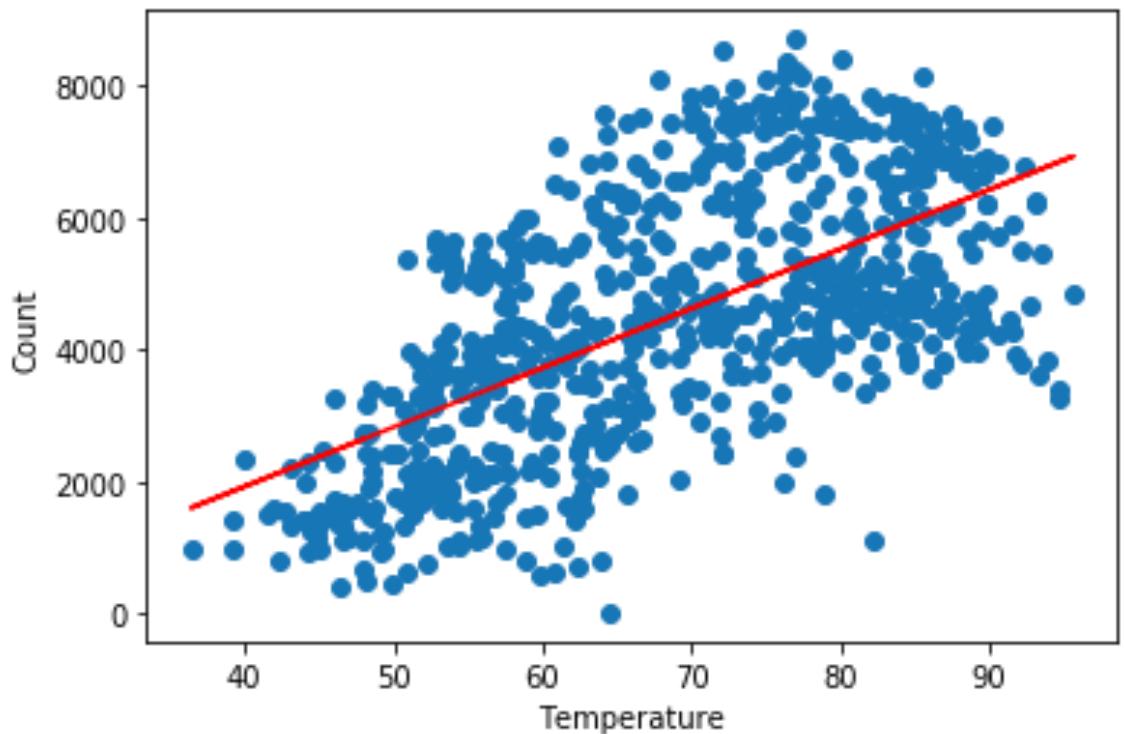
Change in Count by Temperature

- ▶ Clear connection between the change in temperature and count of Bikes
- ▶ Run a single Linear Regression to find the best fit trend line between these two variables



Trend line for our Data

- ▶ The slope of the trend line is 90 and the coefficient -1665
- ▶ This means that for every increase in average daily temperature there are 90 more bikes used.
- ▶ Weaknesses:
 - ▶ Upward trajectory
 - ▶ No realistic coefficient
 - ▶ Only 40% of the data fits



Taking it up a notch & Conclusion

- ▶ New variables included(2012,Holiday,Temp2,Hum)
- ▶ Year had the greatest positive effect at 1949.
- ▶ Temperature was second at 90.
- ▶ Humidity and Holiday both had a negative effect 16 and 579
- ▶ We can predict that 73% of the change in Bike usage is attributed to the above factors.