



Demand Forecasting of Divvy Bikes

Srikrishna Venkat



Presentation Outline

1. Inspiration
2. Data Sources
3. Methodology
4. Evaluation
5. Summary



Inspiration

- I commute from the suburbs and see row of Divvy bike riders going to work.
- Chicago has variable weather.
- Do Divvy bike riders choose to ride a bike based on weather patterns?
- If so, can I predict the total demand of Divvys?



Data Sources

1. Wunderground Weather Data

Hourly temperature readings for each day

Temperature, Dew Point, Humidity, Wind Speed, Pressure, Precipitation etc.

2. Divvy Ridership Data

Divvy System Data released every quarter.

Trip ID, Start Time, End Time, Trip Duration etc.



Methodology

1. Merged DataFrame:

- Rows: 2264
- Features: 54
- Target: 1 (riders)

2. Running Regression in StatsModels:

- Kept only the features that had a p-value < 0.05
- Reduced dataframe (2264, 23)



Methodology

3. Scikit-Learn Linear Regression:

- R^2 on training: 0.39
- R^2 on testing: 0.41

4. Scikit-Learn PolynomialFeatures:

- R^2 on training: 0.37
- R^2 on testing: 0.27



Methodology

5. Scikit-Learn LASSO Regression:

- R^2 on training: 0.43
- R^2 on testing: 0.45



Second Best Model

$$\begin{aligned} Y = & 15.8 * \text{Temp} - 7.9 * \text{Humidity} + 13.1 * \text{Windspeed} - 503.6 * \text{Precip} \\ & - 330.1 * \text{wind_direction_SSW} - 247.9 * \text{wind_direction_SW} + \\ & 173.0 * \text{mood_Light_snow} - 268.1 * \text{mood_Cloudy/Windy} + \dots \end{aligned}$$



Summary

Model does a decent job explaining variance in Divvy ridership based on weather.

- Possible that people don't care about weather when they are biking.
- Next steps: Does weather play a role if I only look at Customer or Explore Pass riders only?