# General Cycle

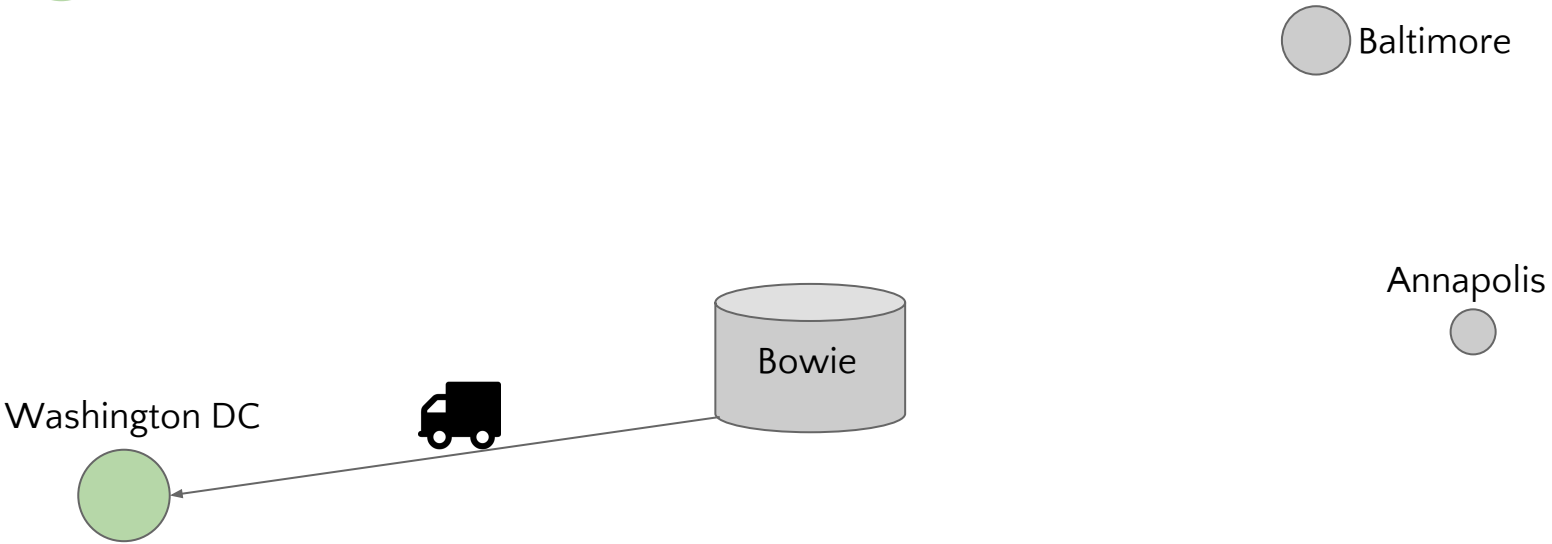
## A bike startup in the DC market

Hussain Burhani

## Lightning points

- Situation
- Business problem
- Feature selection and modeling
- Prediction tool
- Future improvements
- Coda

# Logistics at General Cycle



## 3am at the Bowie facility

- Is it raining today?
- What day of the week is it?
- Is it a holiday?

# Problem statement

***Gather and utilize features which might help predict the number of bike users in any given day in DC.***

**Given a set of weather and time variables, predict the number of bike users in a given day.**

# Linear regression process


- Load dataset
- Ordinal and nominal

- Train / test split
- Feature engineering

- Fit and predict
- Tuning variables
- Evaluate

# Data categories

**Numeric**

Ordinary values

- temp
- temp_feel
- humidity
- windspeed
- *count*

- **As-is**

**Nominal**

Discrete number of categories

- is_holiday
- is_workingday
- weather

- **Dummify**

**Ordinal**

Ordering matters

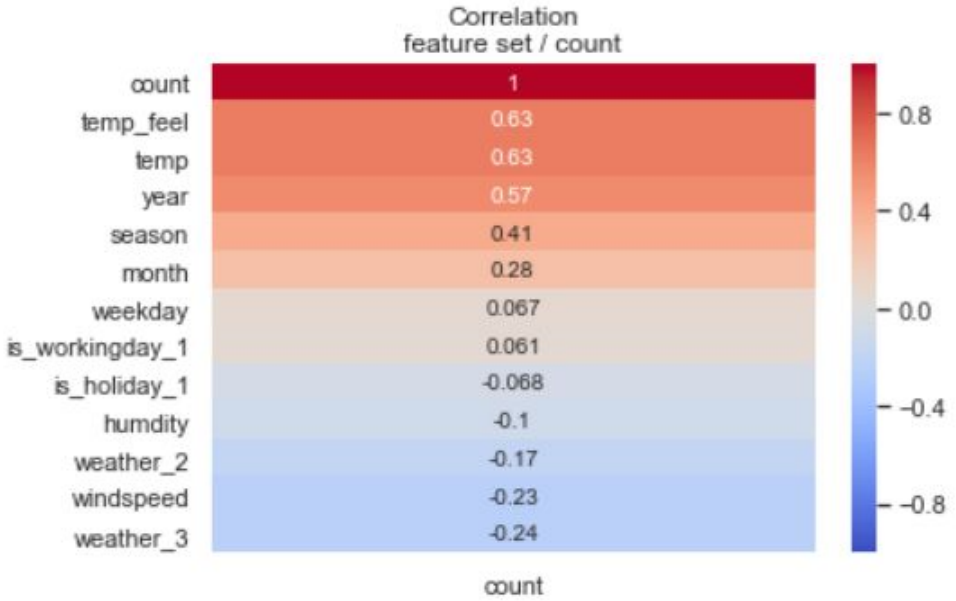
- season
- year
- month
- weekday

- **Ordered numeric map**

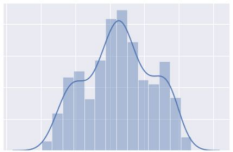
# Correlations of features

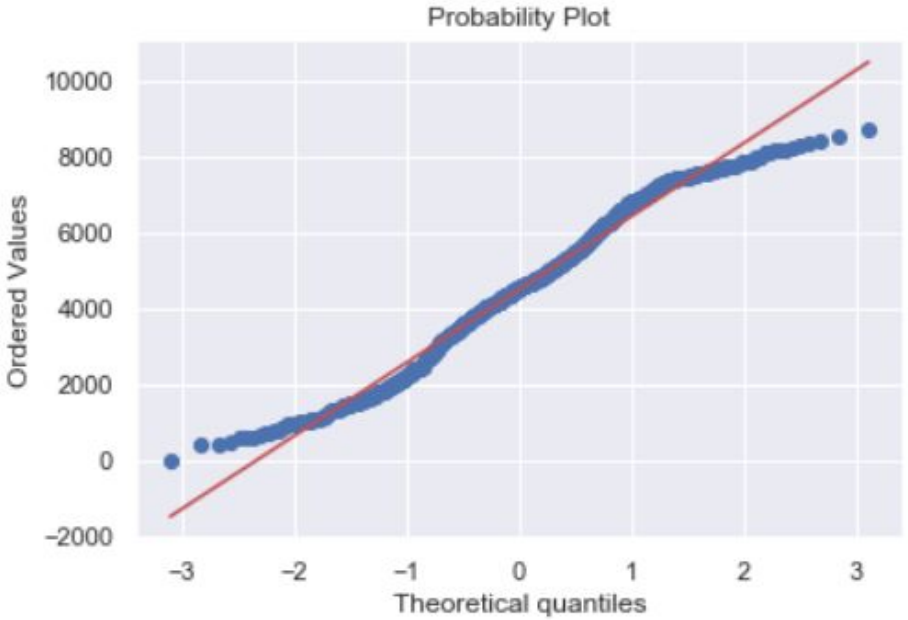
weather (legend):

1. Clear
2. Mist
3. Light Snow
4. Heavy Rain

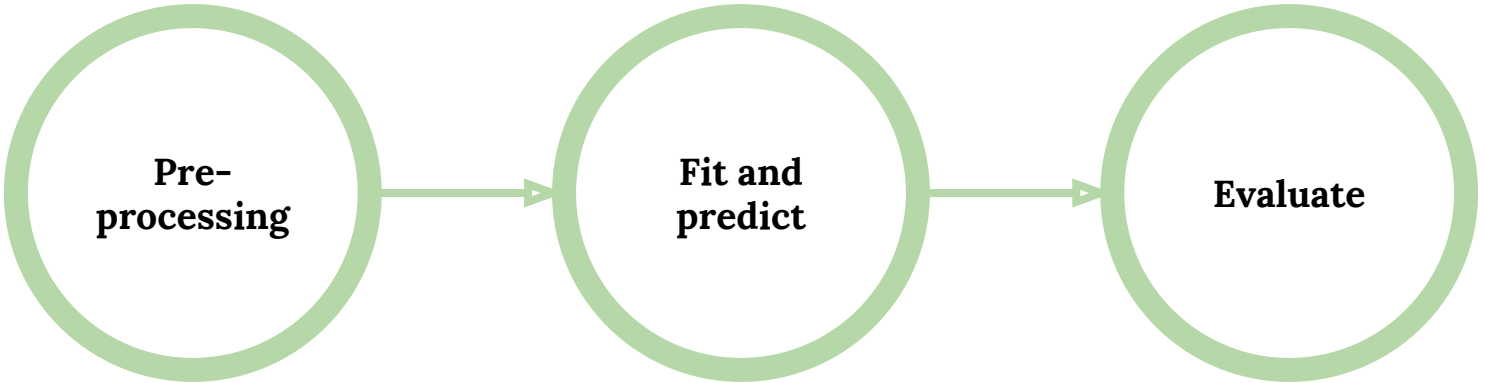


# Normality of target





# Optimal modeling framework (so far)




- StandardScaler
- PolynomialFeatures

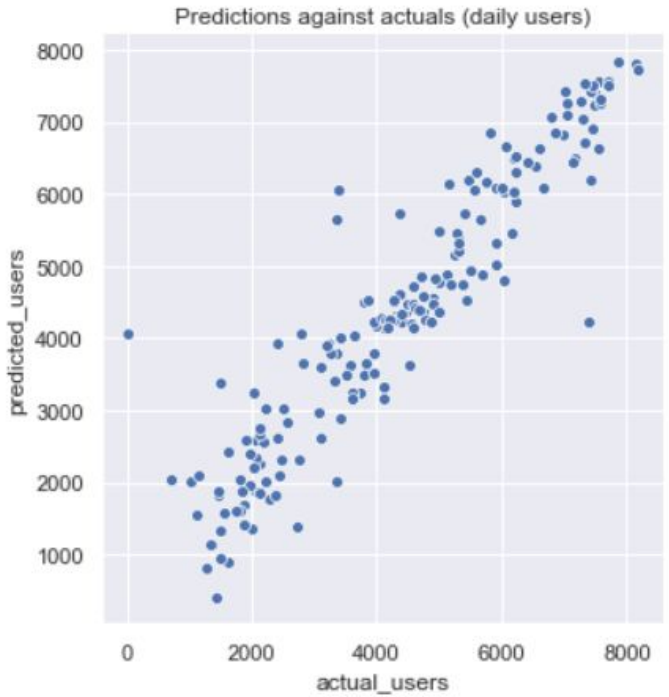
- ElasticNet regularization
- GridSearch

- Compare train/test
- R-squared, RMSE
- Bias-variance

# Predictions

R-squared score:

- training data: 0.9302
- testing data: 0.8705



# Crystal ball for General Cycle

***The operations manager can predict how many additional bikes are needed.***

**Given a set of weather and time variables, predict the number of bike users in a given day.**

# Future improvement

## Features

- Expand feature set
  - Demographics, user type

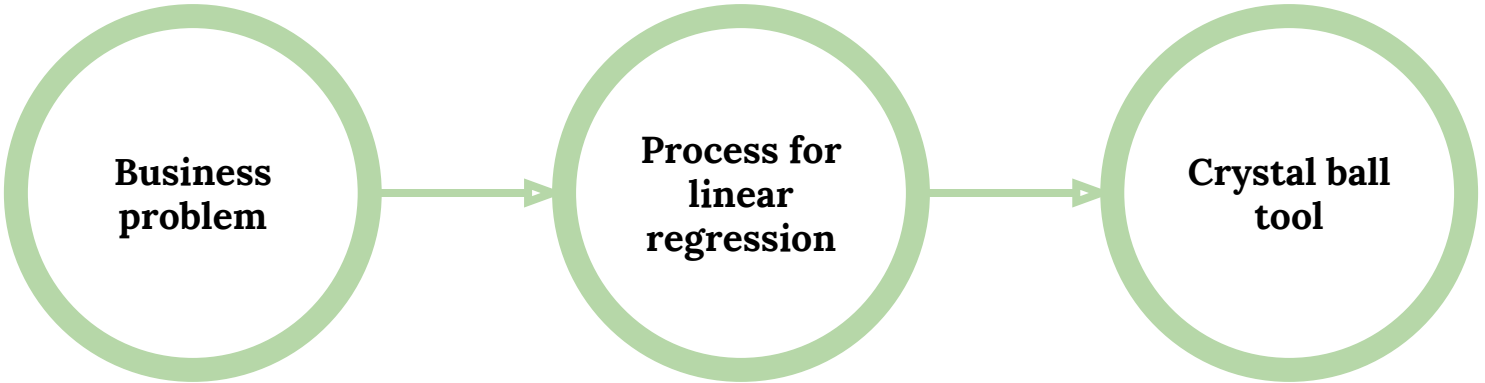
## Modeling

- Alternate modeling frameworks
  - kNN, RandomForest
- Statsmodels for p-values
- Timeseries

## Tool

- Current / live weather info
- Connected to user database
- More features
- Predictions at specific metro stations
- Expanded to other locations

# Coda


Business problem
Process for linear regression
Crystal ball tool