General Cycle

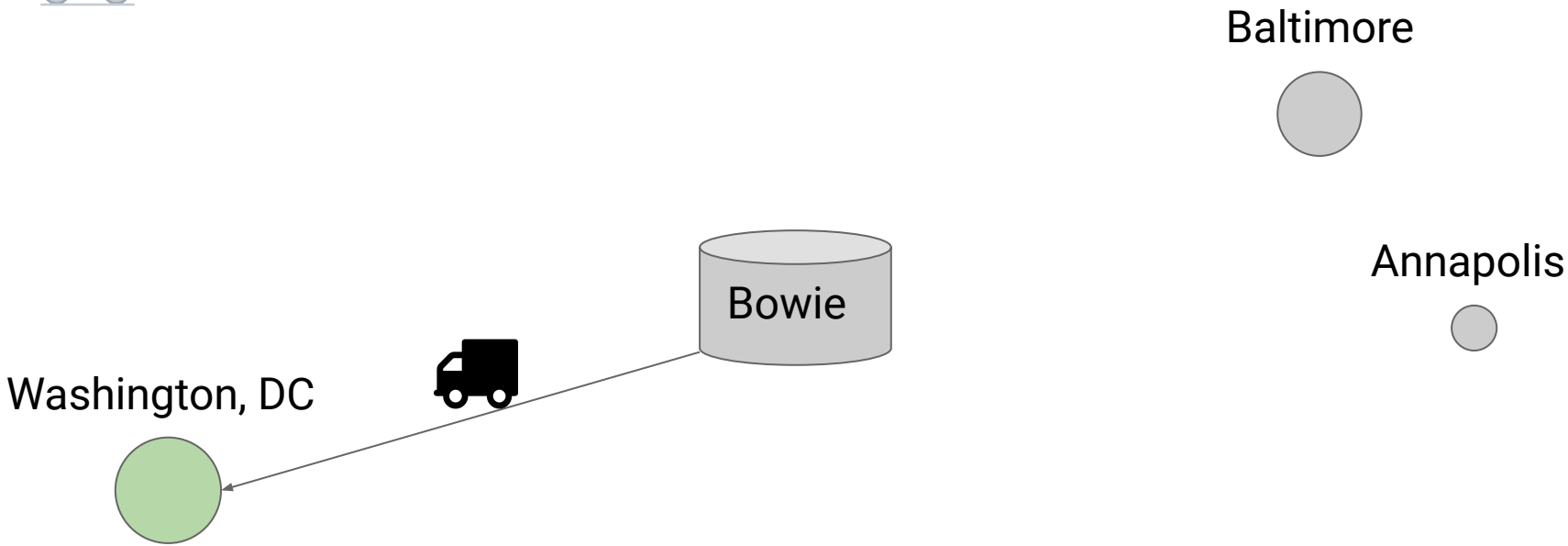
A bike start-up in Washington, DC

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?
- Which day of the week is it?
- Is it a holiday?

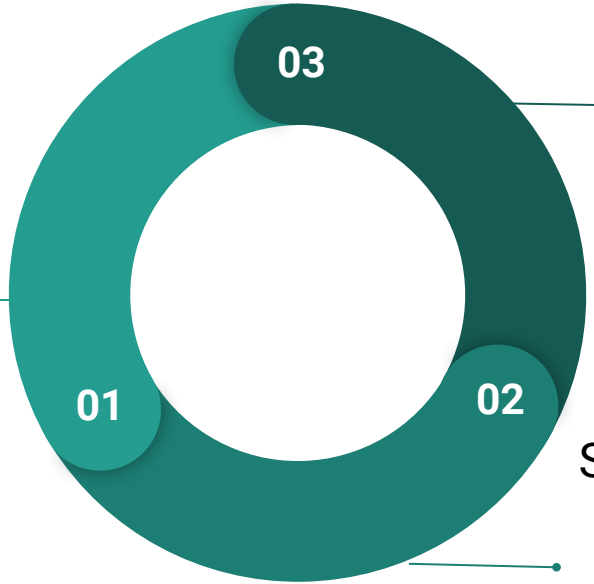
Problem statement

- Business problem: predict the number of bike users in any given day in Washington, DC
- Data science problem: given a set of weather and time variables, predict the number of bike users in a given day

Regression workflow

Gather, clean, and explore data

- Load dataset
- Identify ordinal and nominal



Build, evaluate, improve

- Fit and predict
- Tune parameters
- Evaluate

Split data and select features

- Train, test, split
- Engineer features

Data types

Numerical

- temp
- temp_feel
- humidity
- wind_speed
- ***As-is***

Nominal

Categorical

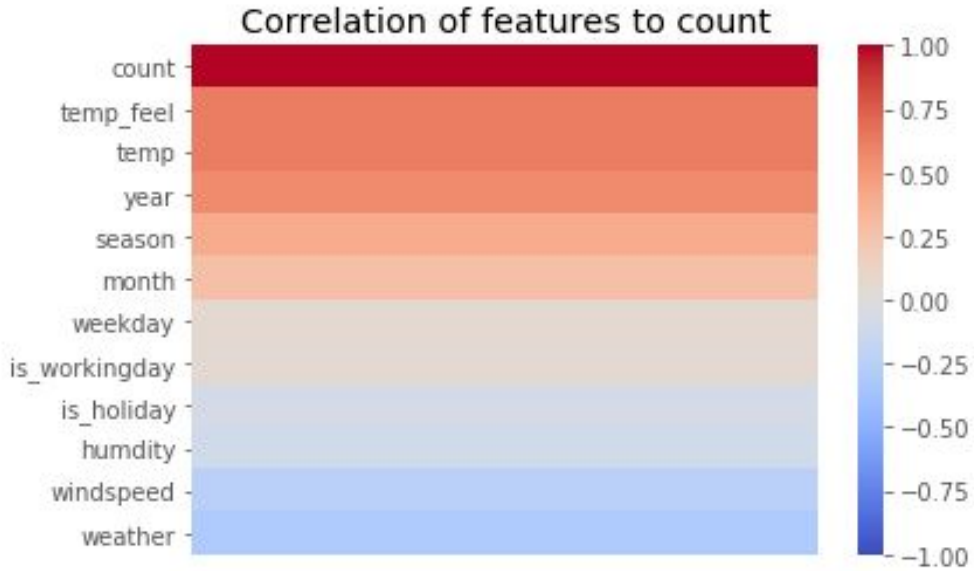
- is_holiday
- is_workingday
- weather
- ***Dummy-ify***

Ordinal

Categorical where order has value

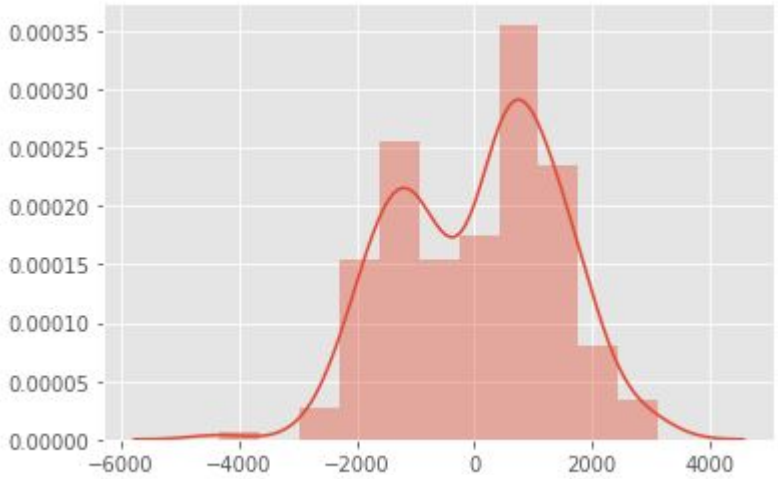
- <none>
- ***Ordered numerical map***

Correlation of features to target



weather (legend)
1: Clear
2: Mist
3: Light snow
4: Heavy rain

Histogram of errors and evaluation



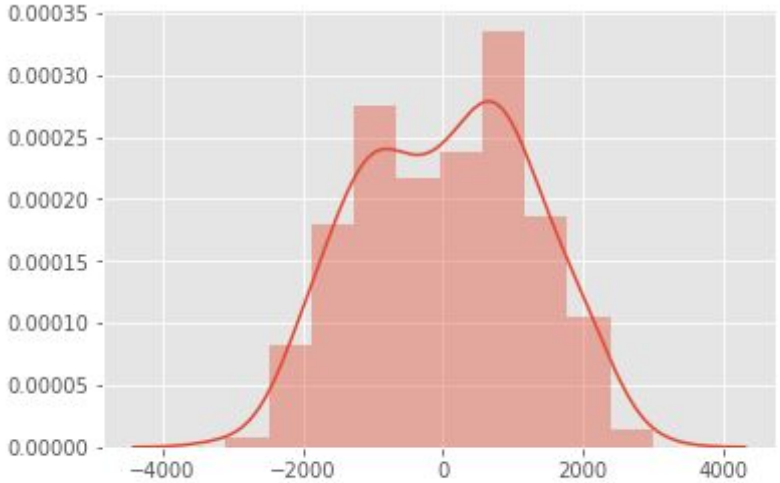
Linear regression with regularization

R-squared

Train: 0.5825

Test: 0.5511

Bias-variance ratio: 0.9461



Catboost

R-squared

Train: 0.9427

Test: 0.6257

Bias-variance ratio: 0.6638

Prediction tool

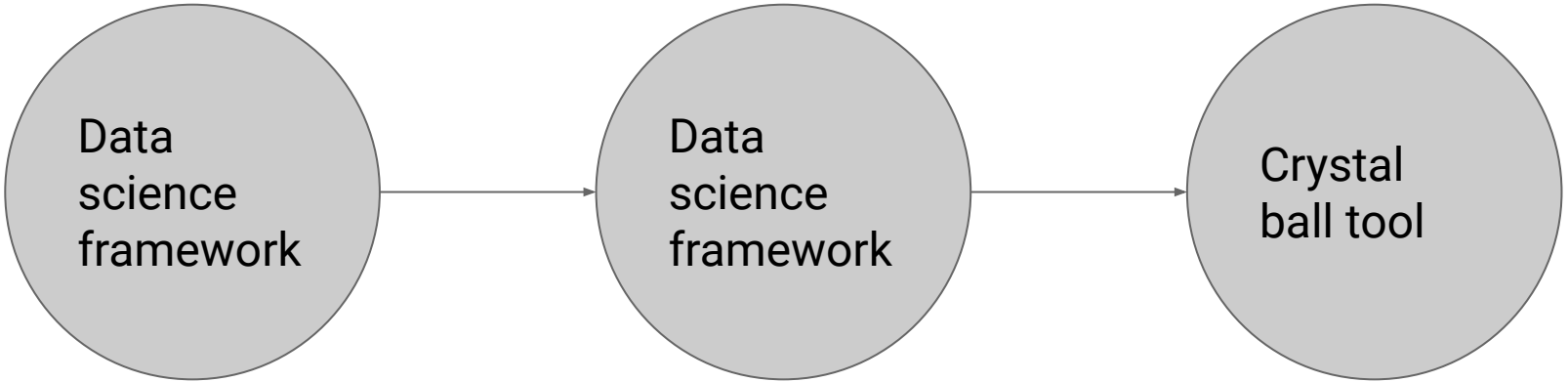
- The operations manager can predict how many additional bikes are needed by providing a set of time and weather variables

- Prediction tool

Future improvements

- Data
  - More observations
  - Expanded feature set (demographics, user_type)
  - Transfer learning
- Modeling
  - Ensemble models
  - Deep learning (with expanded dataset and features)
- Tool
  - Flask-based and hosted app

Coda


Data science framework
Data science framework
Crystal ball tool