Introduction to Data Science

Exploratory Data Analysis (EDA)

April 18, 2019

Today's Agenda

I. ETL Recap

II. Introduction to EDA

III. EDA Example in Python

IV. Q&A

ETL Recap

- Data Sources: SQL, NoSQL, APIs, Flat Files
- Data Structures: Tabular, JSON
- *Python Libs:*   *SQLAlchemy, PyMongo, Requests, Pandas*

Introduction to EDA

- Data Types: Numerical, Text, Boolean, Categorical, Ordinal, NaN
- Summary Statistics: Count, Mean, Med, Min, Max, STDEV, IQR
- Visualizations: Histogram, Boxplot, Probability Density, Heatmap
- *Python Libs:*   *Pandas, NumPy, SciPy, Matplotlib, Seaborn*

EDA Example in Python

Q&A