

# Data Science Final Project

## Lightning Talk

Rafael Santana

May 2017

General Assembly - Santa Monica

# Problem 1 - Predict the best connection time for your trip

Problem statement:

- Based on the airport/airline/date/time find the best connection time to avoid delays.

Data:

- 2015 Flight Delays and Cancellations from DOT (Kaggle)

Hypotheses:

- Train a model that can predict the probability of a delay and the delay time (if exists) to suggest you the best connection time you need to do you trip safely.

# Problem 2 - Find suspicious payments from Brazil's House of Representatives

Problem statement:

- Identify fraudulent or misconduct payments from politicians using public money.

Data:

- Full spend reports
- Don't have fraudulent x no-fraudulent classification

Hypotheses:

- Use clustering and logistic regression to identify fraudulent x non-fraudulent payments

# Problem 3 - Identify correlation between salary and job satisfaction

Problem statement:

- How much salary influences the job satisfaction feeling and what other variables does contribute to a higher job satisfaction?

Data:

- IBM HR Analytics Employment Attrition & Performance from Kaggle.com

Hypotheses:

- Use regression models to identify the correlation between the variables