Project Presentation
Taxis to Revenues

By Connor Van Cleave, Conor Oliver & Teddy Owen

Introduction

❖ Investment Banking Analysts and Associates work late nights

❖ More deals = later nights = more revenue

❖ Taxi drop offs and pickups data available by coordinates

❖ Taxi data around firms = proxy for how many late nights = predictor of future revenues?

Data Layout

❖ Segment Revenues by Firm:
  ➢ Interested in investment banking revenue
❖ Firm Coordinates
  ➢ 4-point polygon outlining pickup zone for each firm
❖ Yellow Taxi Pickups
  ➢ Coordinates of location
  ➢ Uber Data not yet available
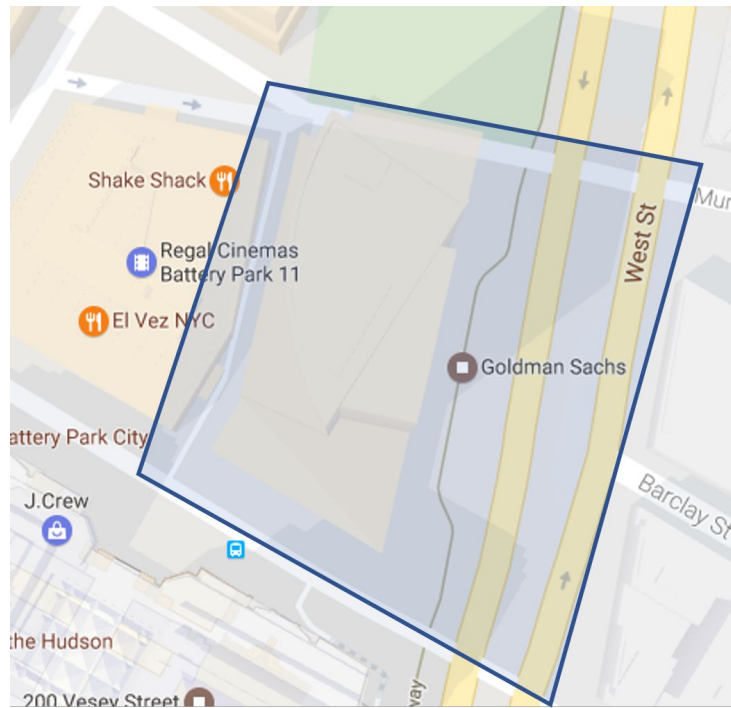
Error Term

- ❖ Unobserved variables that correlate w/ taxi data and affect revenue
  - ➢ Firm crises -> late hours, but reduces revenue
- ❖ Non-employee taxi data
  - ➢ Noisy x variable
  - ➢ Non-banking employee rides
- ❖ Possible that employees work similar hours on deals with lower revenue as those with high revenue
- ❖ Firm-specific effects
  - ➢ Should observe each firm separately
- ❖ Use of other ride services not available

Thus Far

❖ Collected data
- ➢ Segment revenues by quarter: Bloomberg
- ➢ Coordinates for Firms: Google Maps
- ➢ Taxi Data: NYC Taxi and Limousine Commission API

❖ Filtered and Cleaned Data
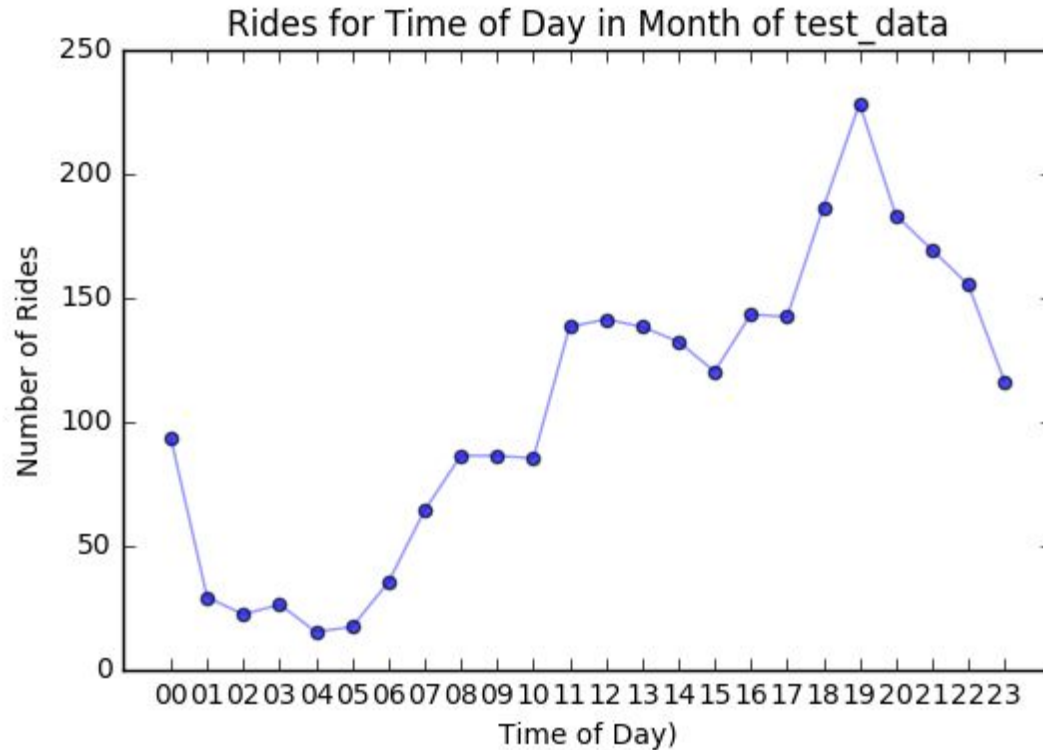- ➢ Isolated taxi data to firm locations

❖ Visualizations
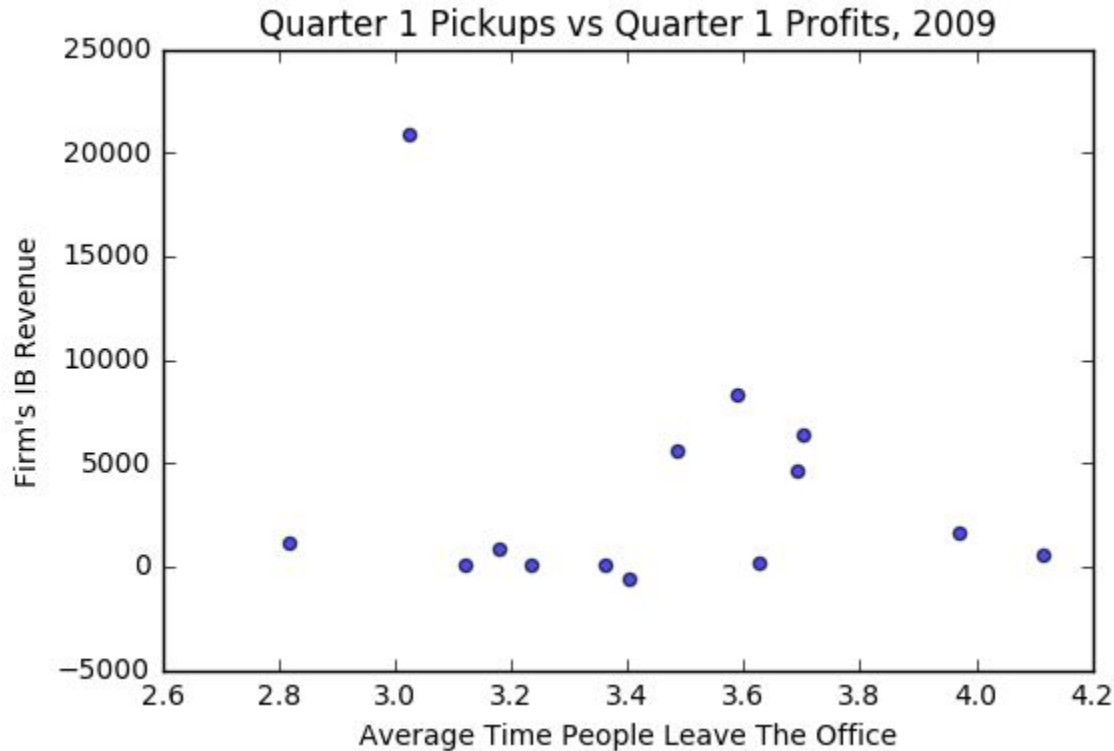- ➢ Line graph, scatter plot

❖ Model Progress
- ➢ Scatter plot
- ➢ Simple Linear Regression
- ➢ IB Revenue Q_{n+lag} = β_0 + β_1 IB Hours Worked Q_n + ε
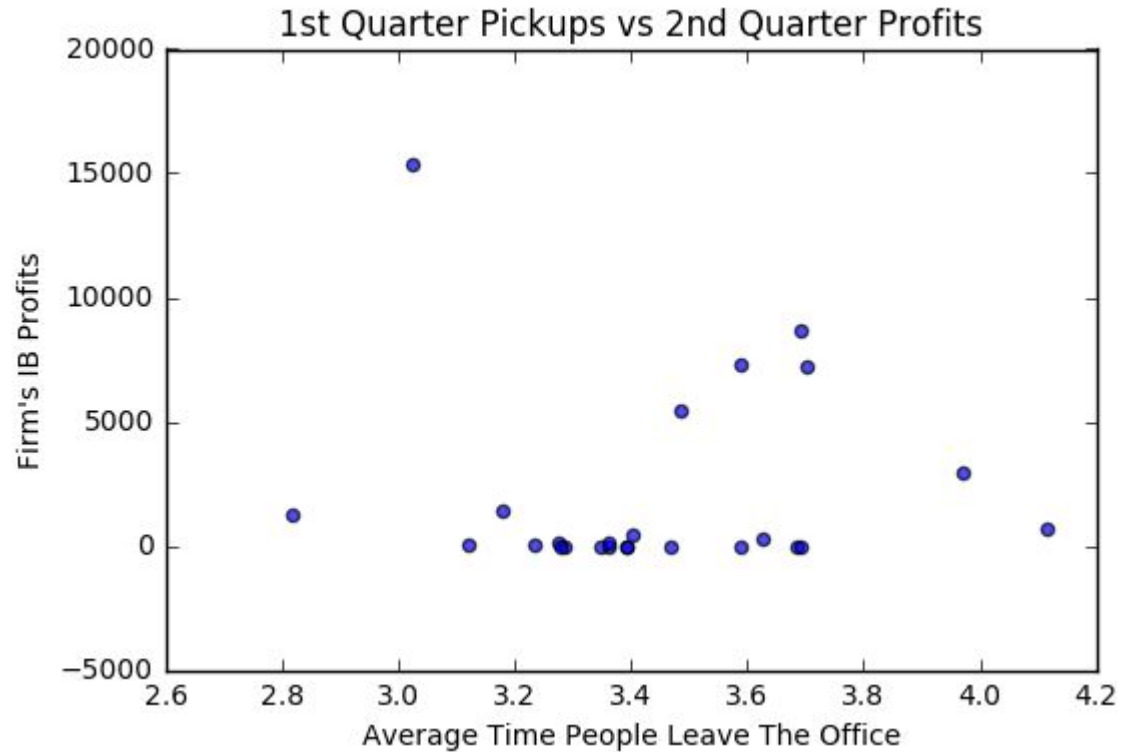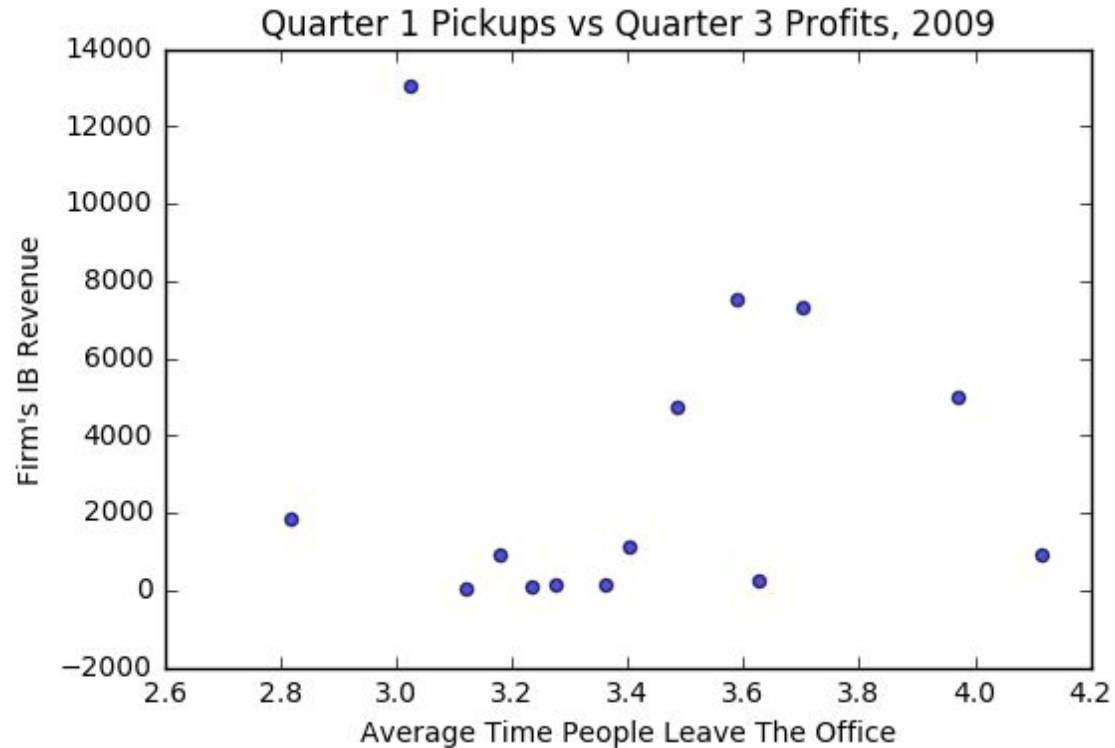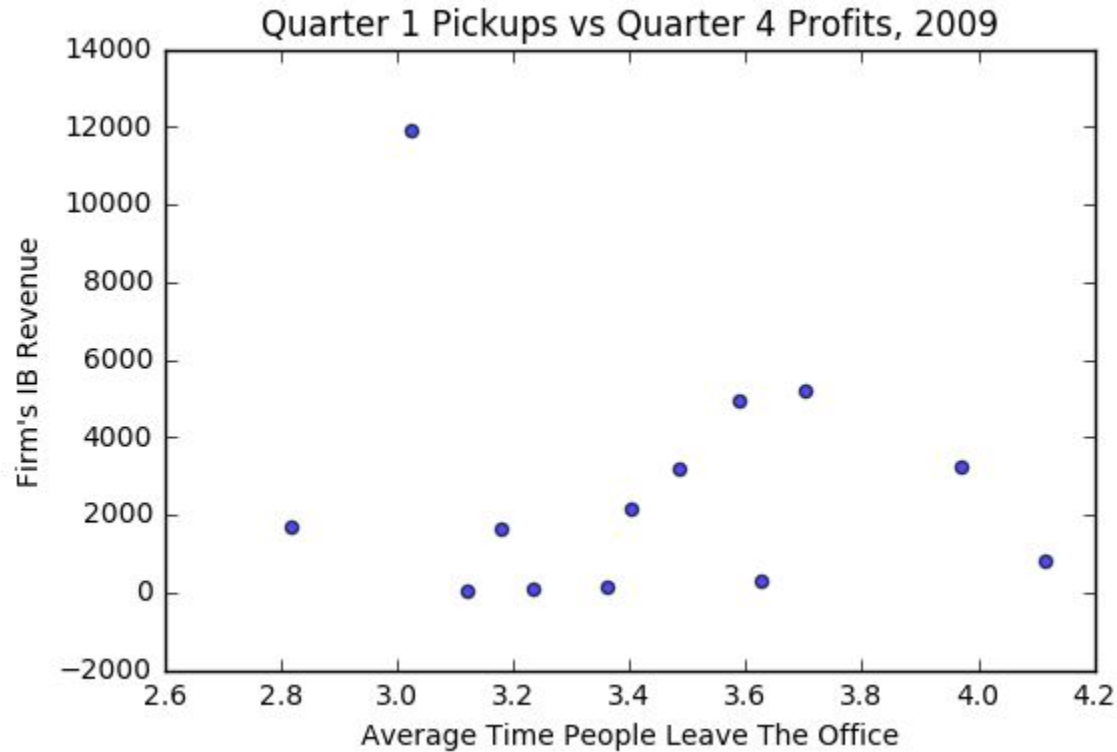- ➢ IB Hours Worked proxy = avg. daily lateness

Visualizations

Rides for Time of Day in Month of test_data

Visualizations

Quarter 1 Pickups vs Quarter 1 Profits, 2009

Visualizations

1st Quarter Pickups vs 2nd Quarter Profits

Visualizations

Quarter 1 Pickups vs Quarter 3 Profits, 2009

Firm's IB Revenue vs Average Time People Leave The Office

# Visualizations

Quarter 1 Pickups vs Quarter 4 Profits, 2009

Elements of Data Science

- ❖ Data collection and processing
  - ➢ Pull GPS coordinates, financials from Bloomberg and using Taxi data API
  - ➢ Linking in SQL and querying
- ❖ Data exploration
  - ➢ Stats and Visualization
- ❖ Data analysis: statistics, hypothesis testing, linear modeling
  - ➢ Simple Linear Regression
  - ➢ Hypothesis Testing (T-tests and F-tests)
- ❖ Machine learning: prediction, clustering
  - ➢ Using Simple Linear Regression for prediction
- ❖ Visualization
  - ➢ Regression results, observing departure times across firms etc
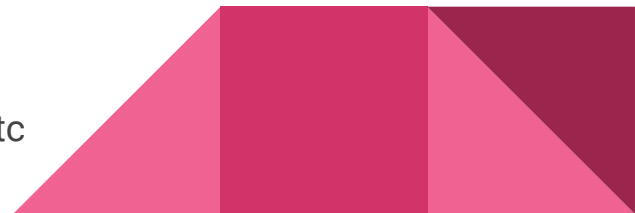
Critical Tasks

❖ Load data into SQL database
❖ Run regressions on many different subsets of data
  ➢ Separately by firm
  ➢ For each segment
  ➢ For total revenue
  ➢ Without 2009
  ➢ With absolute, not average, hours
❖ Process more years of our data