AirBnB Pricing POC

Capstone Project
2/2/18
Sankarshan Archarya, Paul de Fusco, Parinaz Azdavari

Agenda

- Problem Definition
    - Data Science Challenge
    - Datasets: core, Real Estate, Customer, More
    - Data Integration: Data Architecture, ETL/Pipelines
    - Project Status

Problem Definition

- Provide a Pricing tool for AirBnB host that can make effective recommendations to set prices for property managed for each calendar day

Challenge

- Analyze price conditions to determine price recommendation

- Partial information is provided by AirBnB

- No customer data is available - only ‘supply side’ Airbnb data and other real estate datasets

Core Dataset

- Inside Airbnb: http://insideairbnb.com/get-the-data.html
  - Listings, calendar, reviews
  - Will scrape more similar to this

Additional Real Estate Datasets

- Zillow Real Estate: https://www.kaggle.com/c/zillow-prize-1/data

- Hotel Pricing Data: https://www.kaggle.com/c/expedia-hotel-recommendations/data

Customer Datasets

- What are some things we can learn about customers?
    - Social network data

Additional Factors Datasets

- Weather data?

- Any environmental factors that will affect demand/supply

Data Architecture

Predictive Models

Central Storage (AWS)

InsideAirbnb Datasets

Hotel Data

Zillow Rental & Home Data

Other datasets (weather, consumer, etc)

Project Status

- Goal: run POC with simplified version of the above
- Datasets: core data + sample external data
- Data Wrangling: selected basic feature set (98 cols) by merging the core datasets into a Master Dataset
- Data Exploration: plotted price as function of single variables
- Analytics: executed simple regression models on core datasets