



CIS 4560 Final Project

Annie Chen
Mariana Curiel
Monique Duong
Usmon Muslimbekov

Background



Online
Marketplace

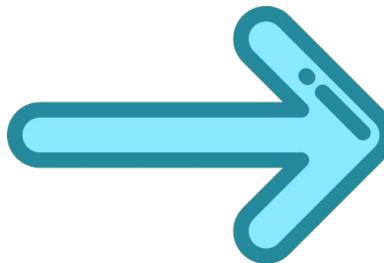
Worldwide
Availability

Founded
in 2008

Introduction (Data Background)

- According to [Inside Airbnb](#), the data is sourced from publicly available information from the Airbnb site
- According to the site, the data has been analyzed, cleansed and aggregated

Introduction (Categorization)



Airbnb
categorizes their
data by county

Then by
neighborhood
directly in the data

Introduction (File Types)



Listings: shows all the airbnb listings in the city/neighborhood.

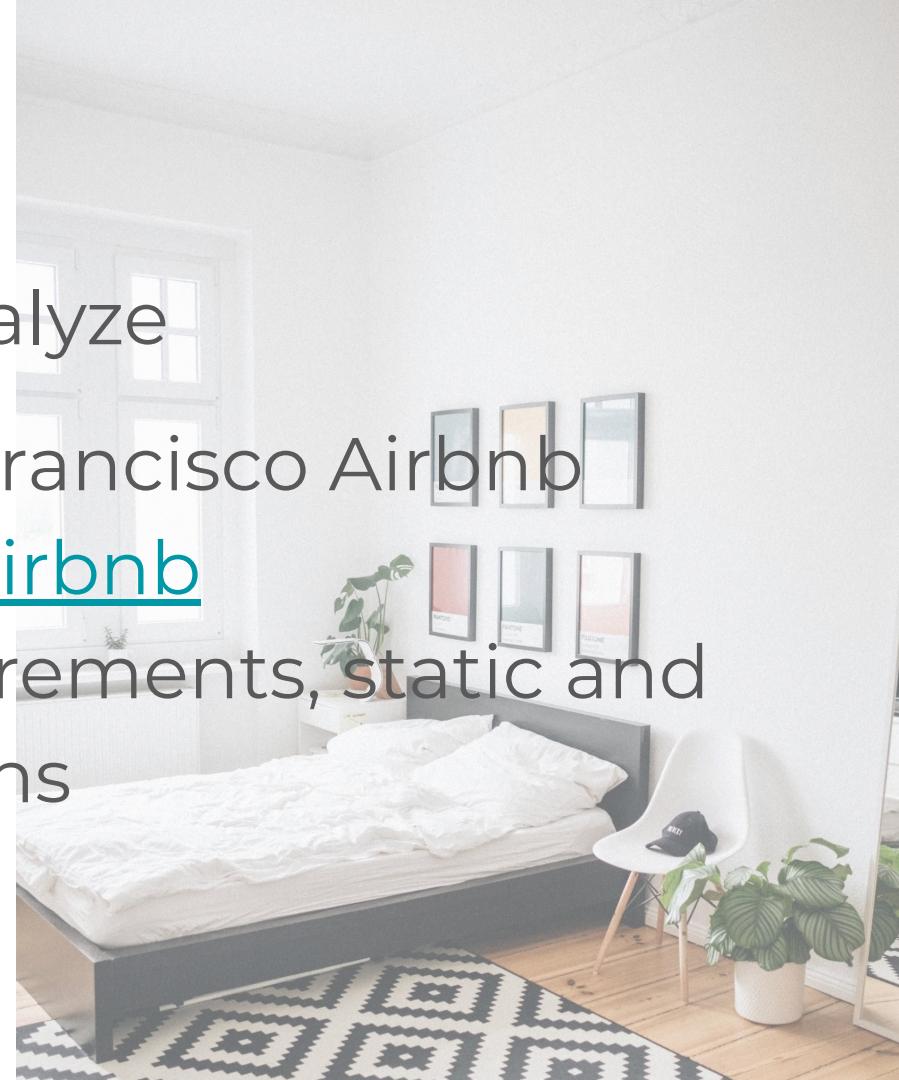


Reviews: shows the reviews for each listing

Abstract

In this project, we will analyze

- Los Angeles and San Francisco Airbnb datasets from [Inside Airbnb](#)
- Using different measurements, static and interactive visualizations



Abstract

- Questions we want to answer
 - Which city is the most popular?
 - Which city is the most expensive? Most affordable?
 - How has Airbnb grown in terms of popularity in Los Angeles?
 - Is there seasonality based on the reviews?

Abstract

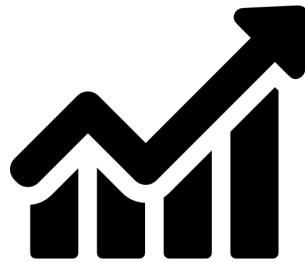
- Measurements we used to gather insight



Analyzing Rental Landscape:
Price by Neighborhood

Abstract

- Measurements we used to gather insight



Analyzing Seasonality:
Number of Reviews by
Quarter, Year



Analyzing Popularity:
Reviews by
Neighborhood

Related Analysis

Boston Airbnb data analysis of 2019



Analyzes Pricing
Trends in properties
by features



Analyzes how reviews
can be predicted
according to different
characteristics

Related Analysis (cont.)

Seattle Airbnb Analysis of 2019



Analyzes trends in rentals by season



Analyzes popularity by types of property and pricing

Description of Data

- LA Listings - key attributes used in the analysis are id(chararray), neighborhood(chararray), latitude(double), longitude(double), price(double)
- LA Reviews - key attributes used in the analysis are listing_id(chararray), date(datetime), reviewer_id(chararray), reviewer_name(chararray)

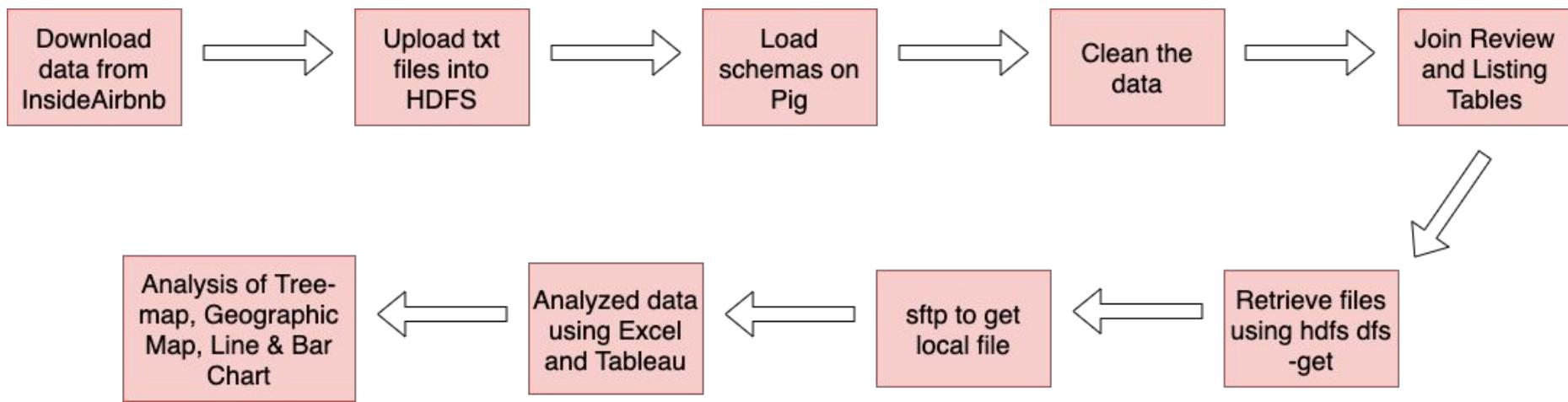
Description of Data

- SF Listings - key attributes used in the analysis are id(chararray), neighborhood(chararray), latitude(double), longitude(double), price(double)
- SF Reviews - key attributes used in the analysis are listing_id(chararray), date(datetime), reviewer_id(chararray), reviewer_name(chararray)

Specifications

- Cluster Version 6.2 (includes Apache Spark 2.4.4, Scala 2.11)
- Cluster number of nodes 1
- Approved Memory size < 2 GB
- Applications: Pig 0.17.0, Excel, Tableau
- CPU speed 2.30 GHz

Flow Chart of Data Analysis

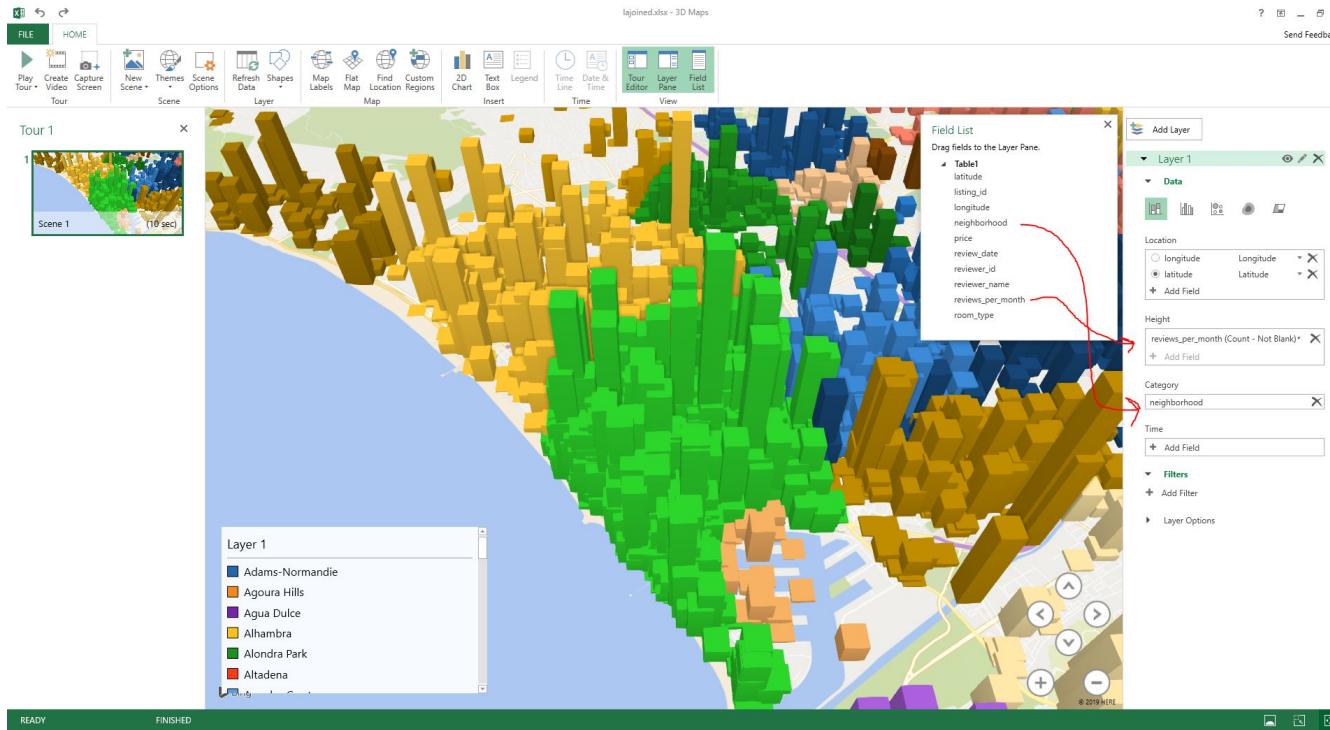


Data Source URL

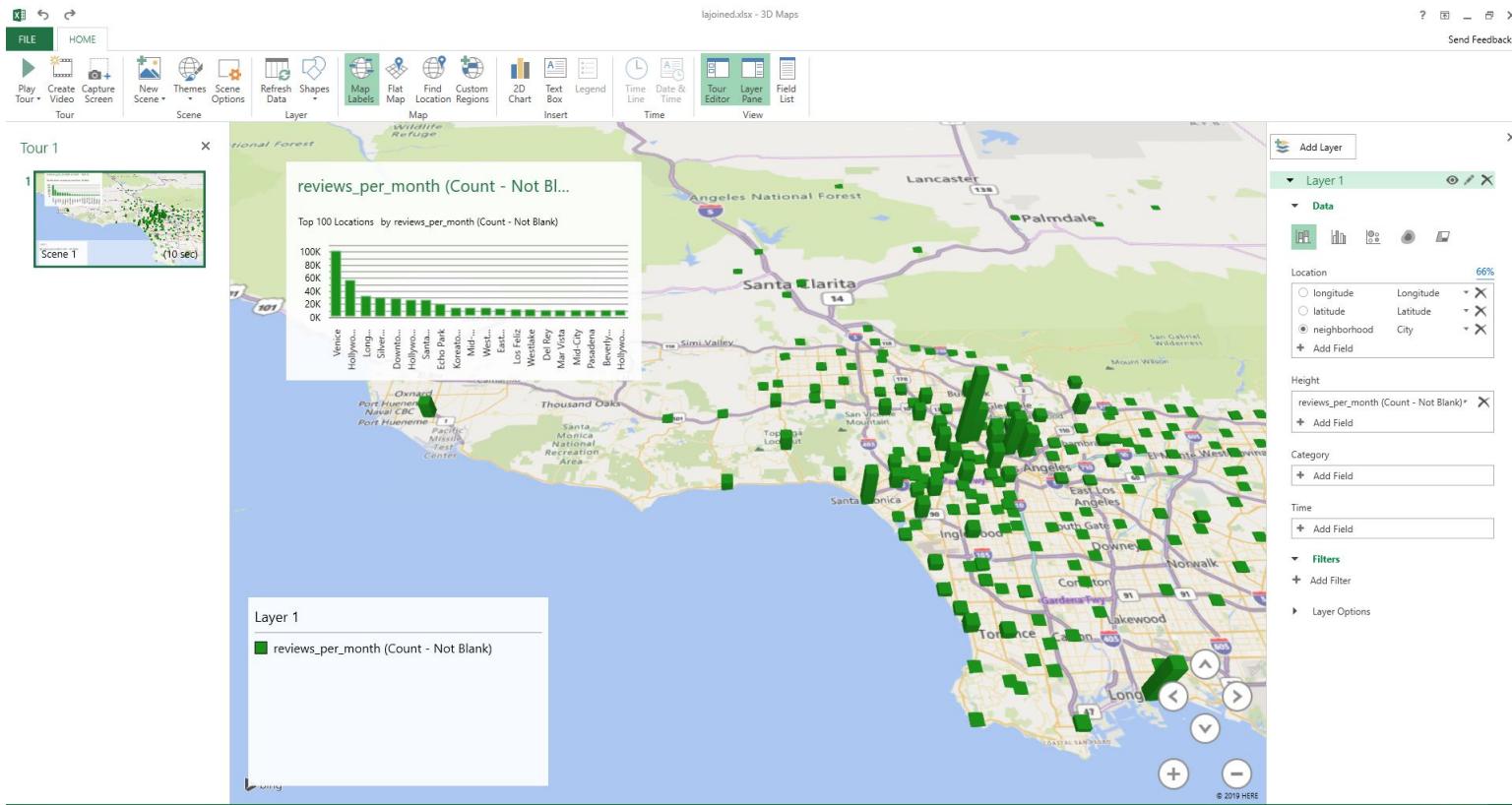
<https://github.com/anniechen61/CIS4560/>

- Presentable slides
- Flow Chart
- Dataset
- Script
- Tutorial

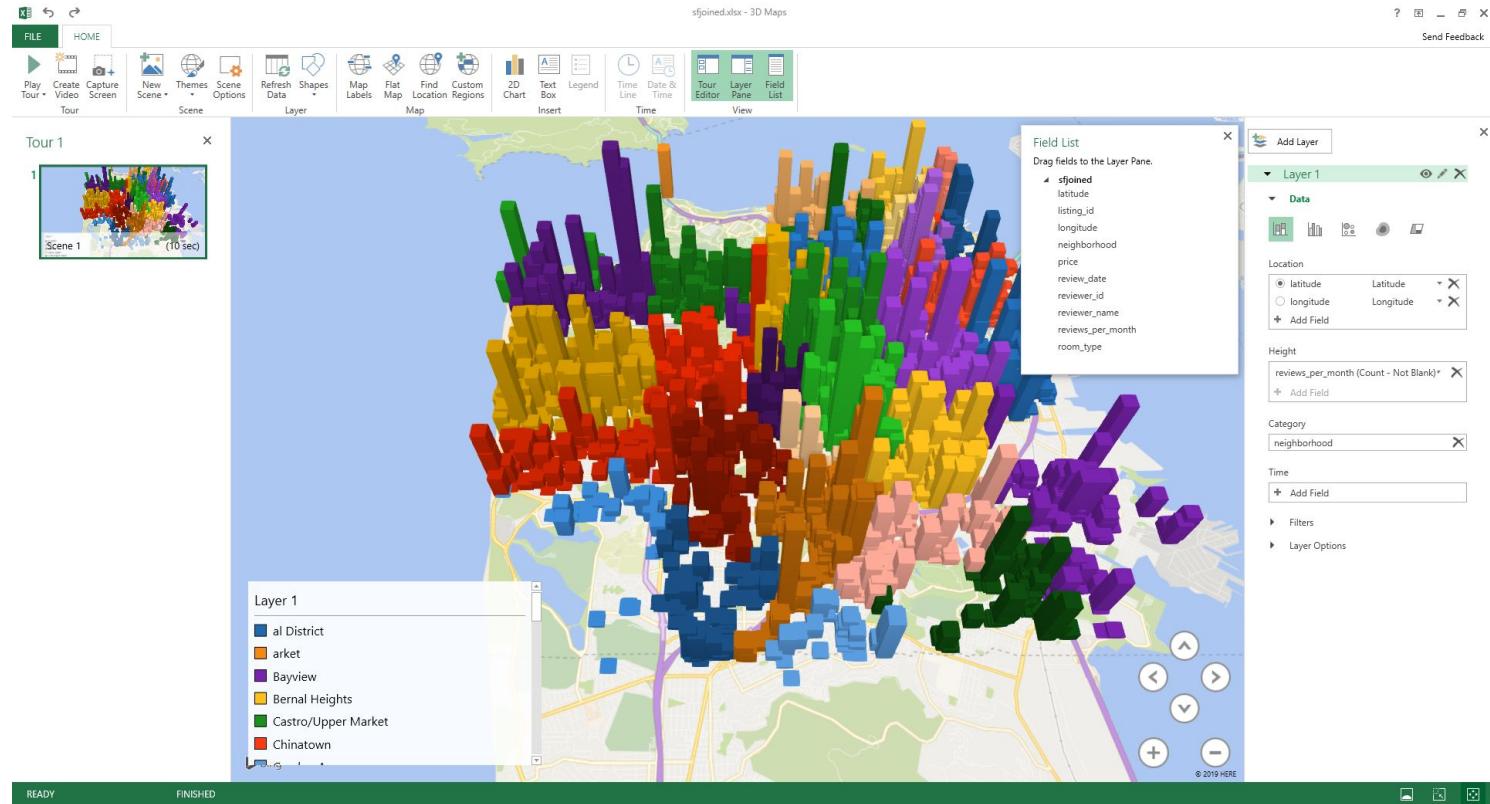
Analyzing LA Data using 3D Maps



Analyzing LA Data using 3D Maps



Analyzing SF Data using 3D Maps



Airbnb Rental in LA

Airbnb Rental in Los Angeles: Neighborhood by Avg Price

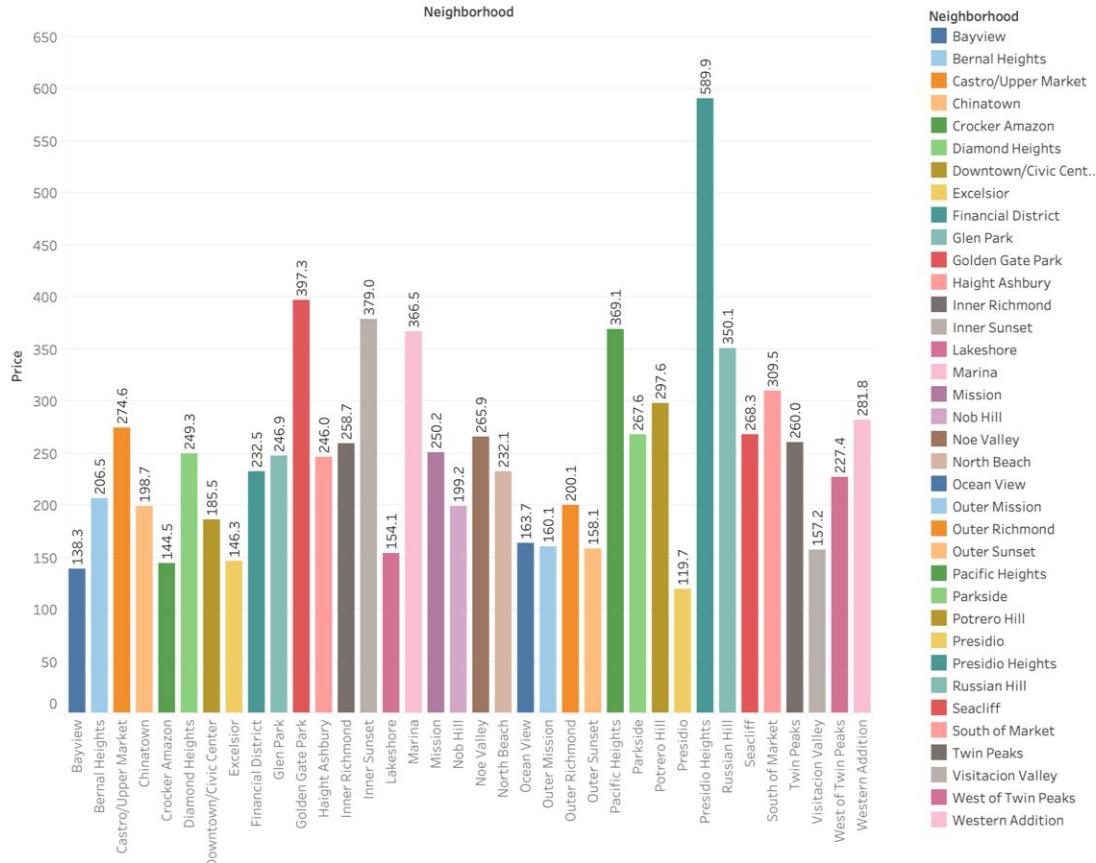
Tree Map showing the most expensive area to the most affordable area to rent Airbnb in Los Angles.



Airbnb Rental in SF

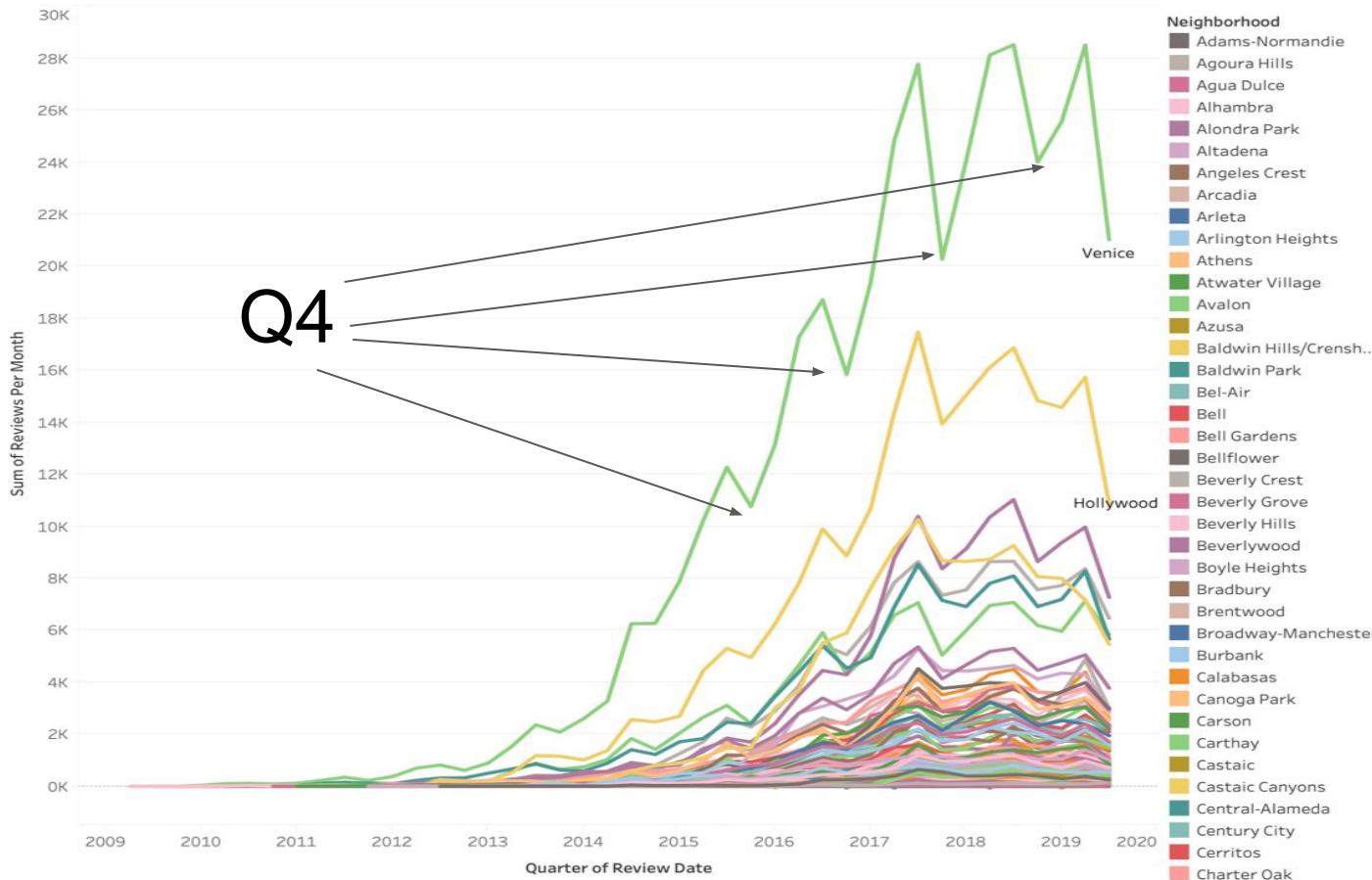
Airbnb Rental in San Francisco: Neighborhood by Avg Price

Bar graph showing the most expensive to the most affordable area to rent Airbnb in SF



LA Trend Analysis 2009-2019

1/1



Things we learned from this project

- Data Engineering phase of the project was difficult
- There were many tools available to us for analyzing the finalized data
- Airbnb has significantly grown and it's apparent by our reviews trend analysis



Thank you!

