

# Estimation of King Estate's Value

Presenter: Jeff Liu





# Factors To Incorporate (King Estate)

## Home Facts:

- Built in 1880 (143 Years)
- Lot Area: 80368 sqft
- Living Area: 9286 sqft
- 8 Bedrooms, 2 Full Baths, 2 Half Baths
- Classification: Residential/Single Family
- Central Heat
- Excellent Condition

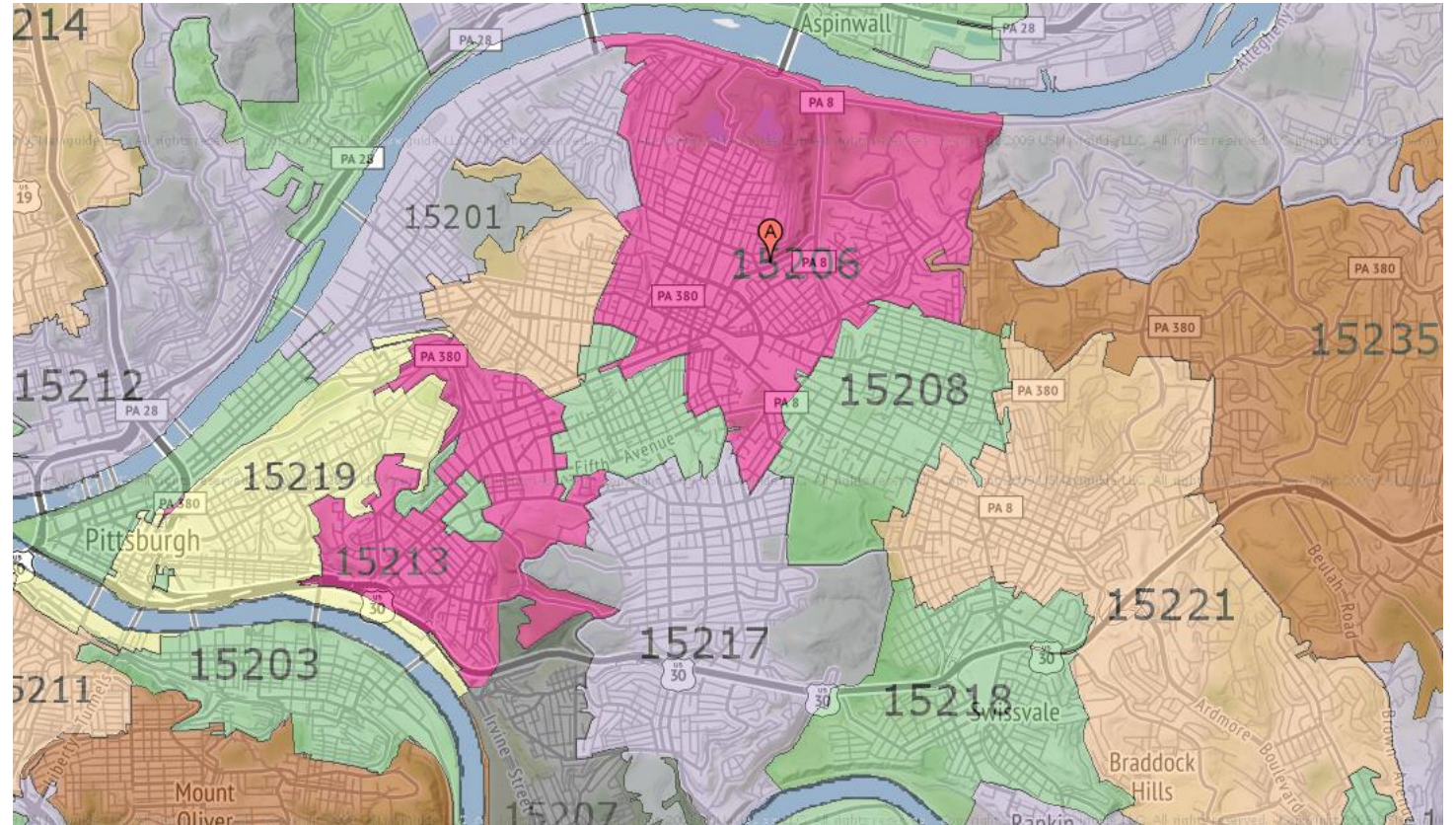


# Factors To Incorporate (King Estate)

---

Location:

- 5501 Elgin St,  
Pittsburgh, PA, 15206
- Highland Park

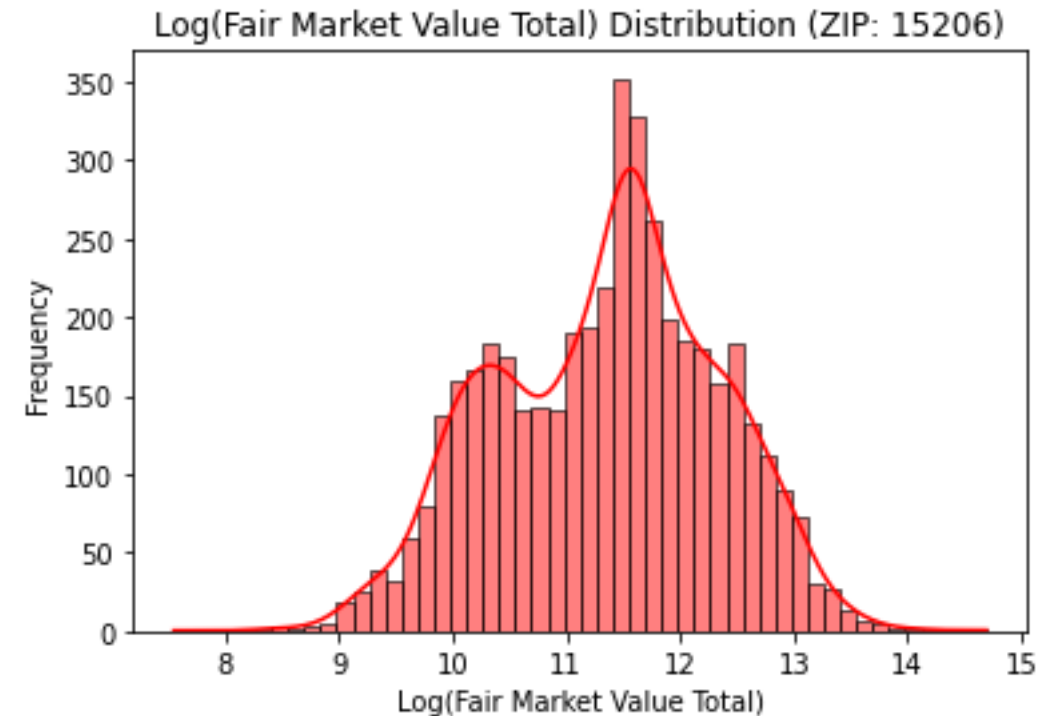
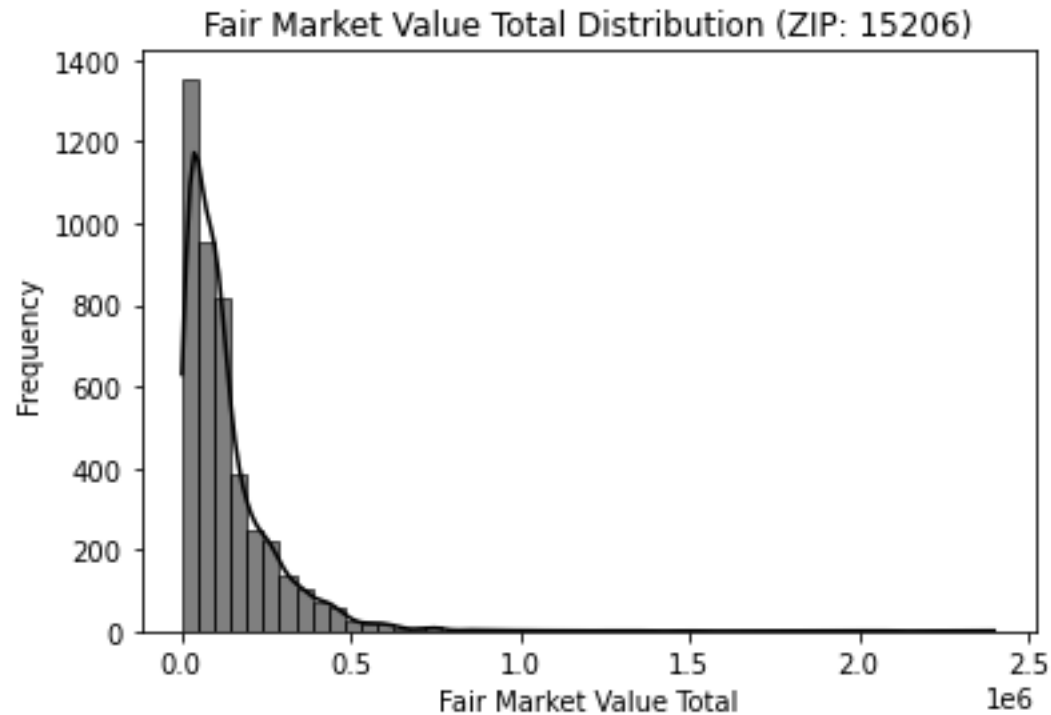


# Data Exploratory Analysis And Cleaning

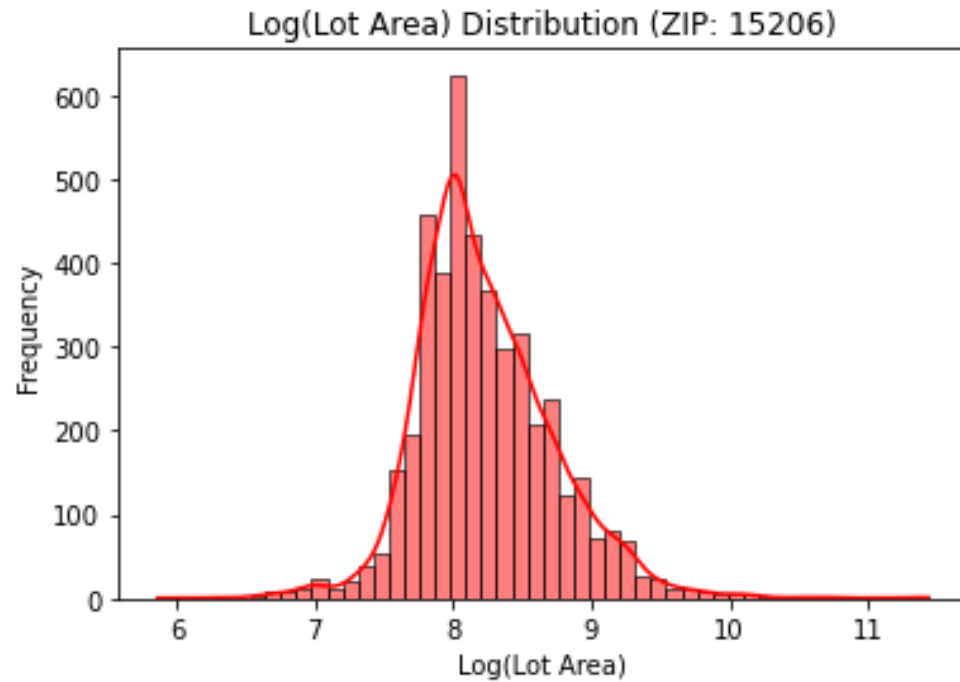
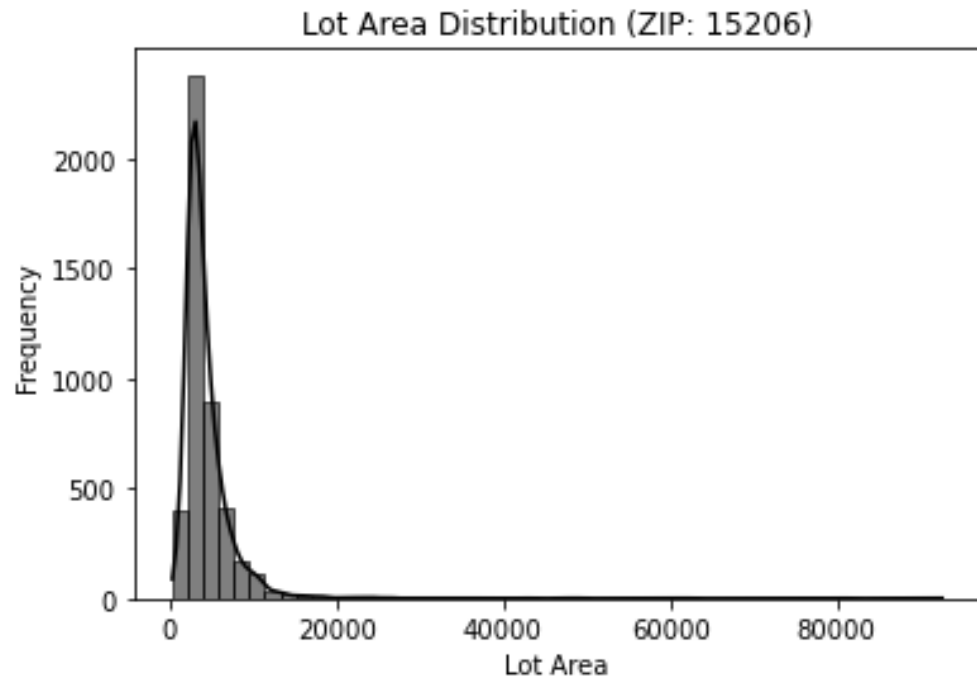
- Visualize distributions and look for outliers
- Filter out houses that are not located at Highland Park (ZIP 15206)
- Only target on residential houses
- Remove year 2020 and onwards due to Covid-19 impact on housing market
- Data size reduced from 580k observations to 4.4k observations



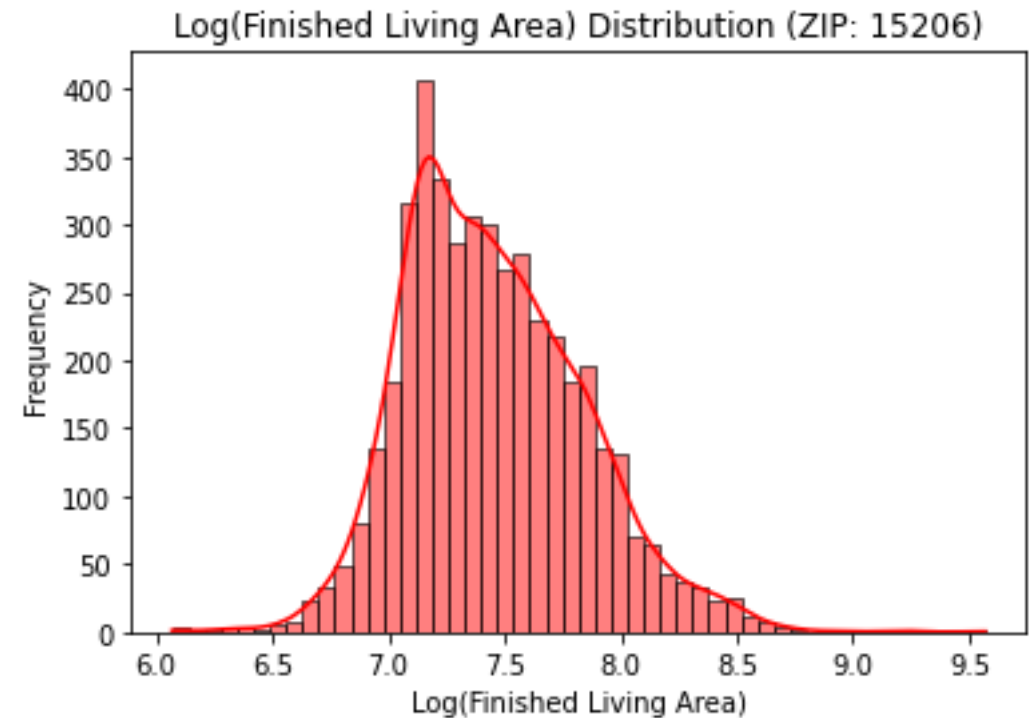
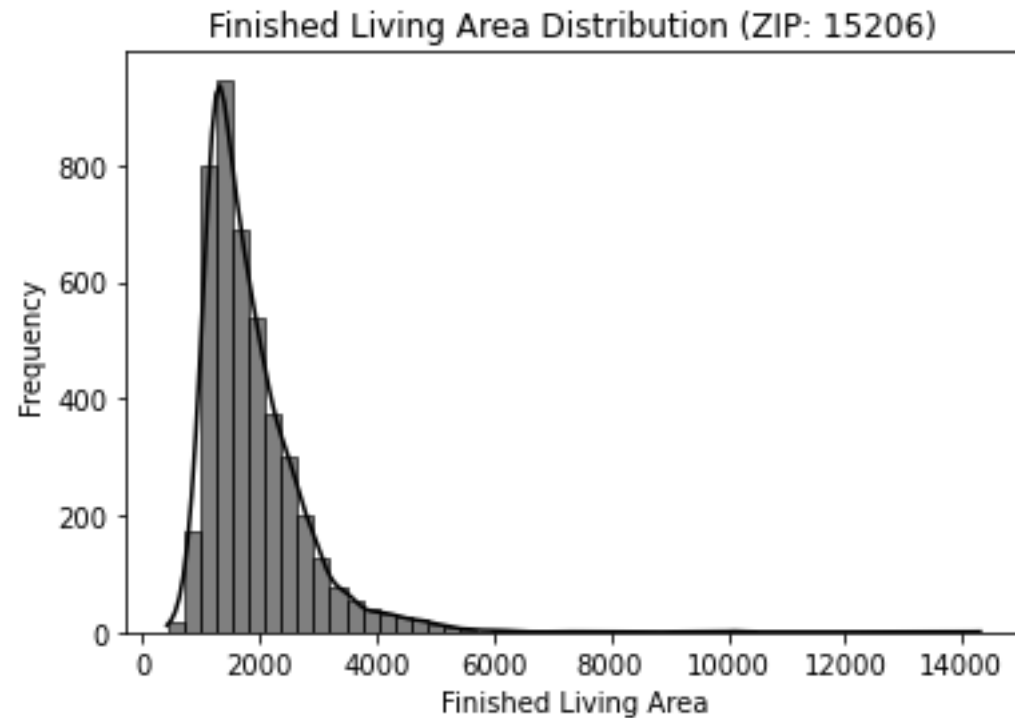
# Feature Engineering: Log Transformation



# Feature Engineering: Log Transformation



# Feature Engineering: Log Transformation



# Approach One: Linear Regression

- Dependent Variable: Log(Fair Market Value Total)
- Independent Variable: Log(Lot Area), Log(Living Area), Building Age, Bedrooms, Full Bathrooms, Half Bathrooms, Fireplaces, Basement Garage
- **ESTIMATED VALUE FOR KING ESTATE: \$1,456,082**
- Note: All coefficients are statistically significant.  $R^2=0.48$



# Approach Two: Gradient Boosting Regression

- Test/Train Split at 20:80 Ratio
- 1000 Estimators (Trees)
- Run model on training dataset
- Predict values on testing dataset
- Model Accuracy: 0.792
- **ESTIMATED VALUE FOR KING ESTATE: \$1,073,178**

# Why housing valuation is difficult?

## Time and uncertainty:

- Unexpected event can happen, which has significant effect on housing market, such as Covid-19 from 2020
- I filtered the dataset, so it only contains observations before 2020.

## Lack of data:

- Although the dataset contains about 600k observations, houses that have similar characteristics like 'King Estate' are **rare**.
- The data may **not be very representative** for King Estate
- I tried to filter the dataset so only similar houses remain in the dataset. However, the tradeoff is a smaller sample size.

# Why housing valuation is difficult?

## King Estate is **special**:

- A luxury and well-maintained property built 143 years ago
- 80368 sqft of lot and 9286 sqft of living area
- Excellent condition
- Houses like 'King Estate' have a smaller market (less buyers)

## Omitted Variables:

- Many, such as intrinsic values, economies, rising interest rates