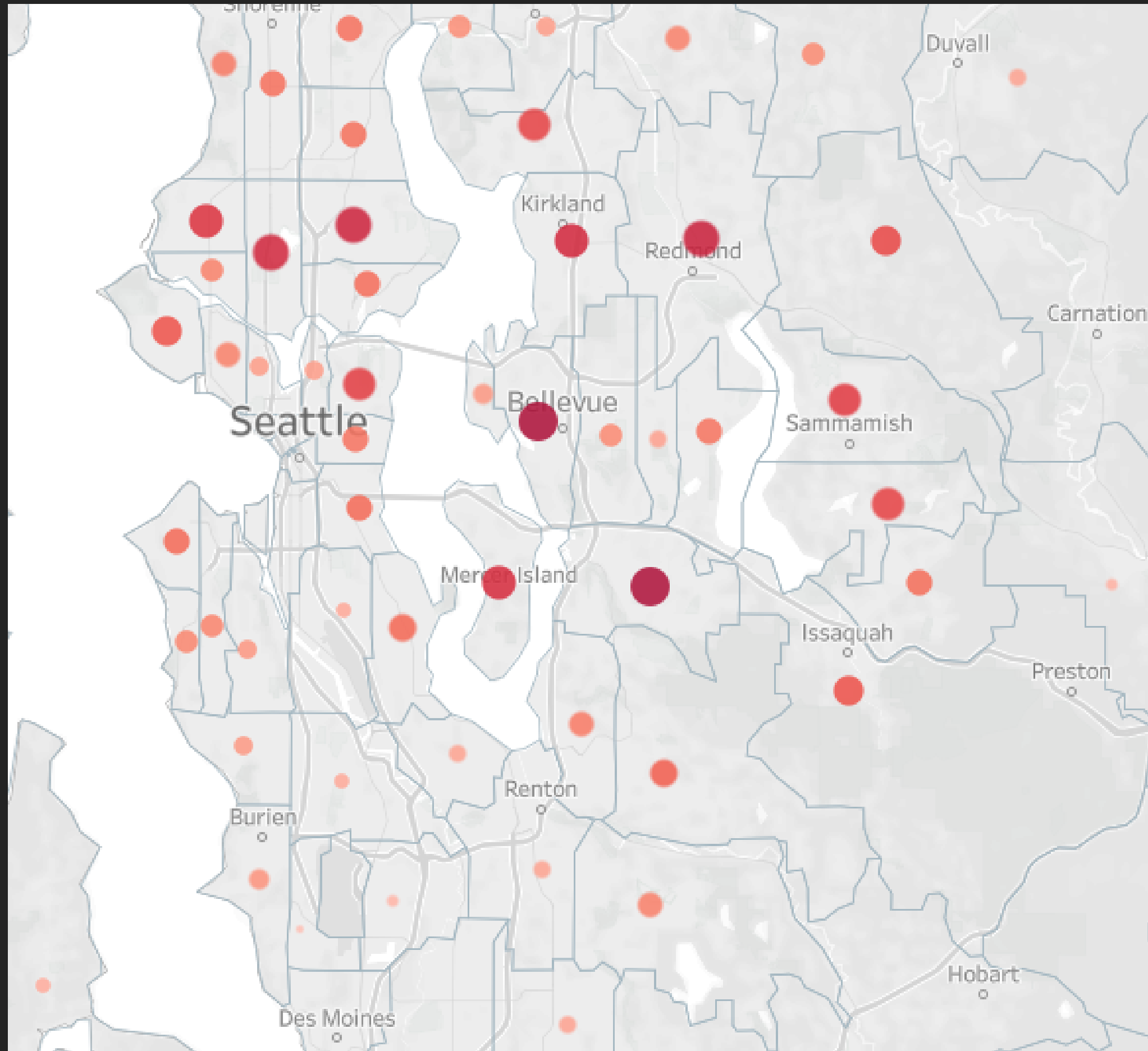# ANALYSIS OF PROPERTIES SOLD IN THE SEATTLE AREA BETWEEN 2014-2015

By Jaan Ugrinsky

Mid-Bootcamp Project
Data Analytics Bootcamp, Ironhack Berlin
(Oct-Dec)2020
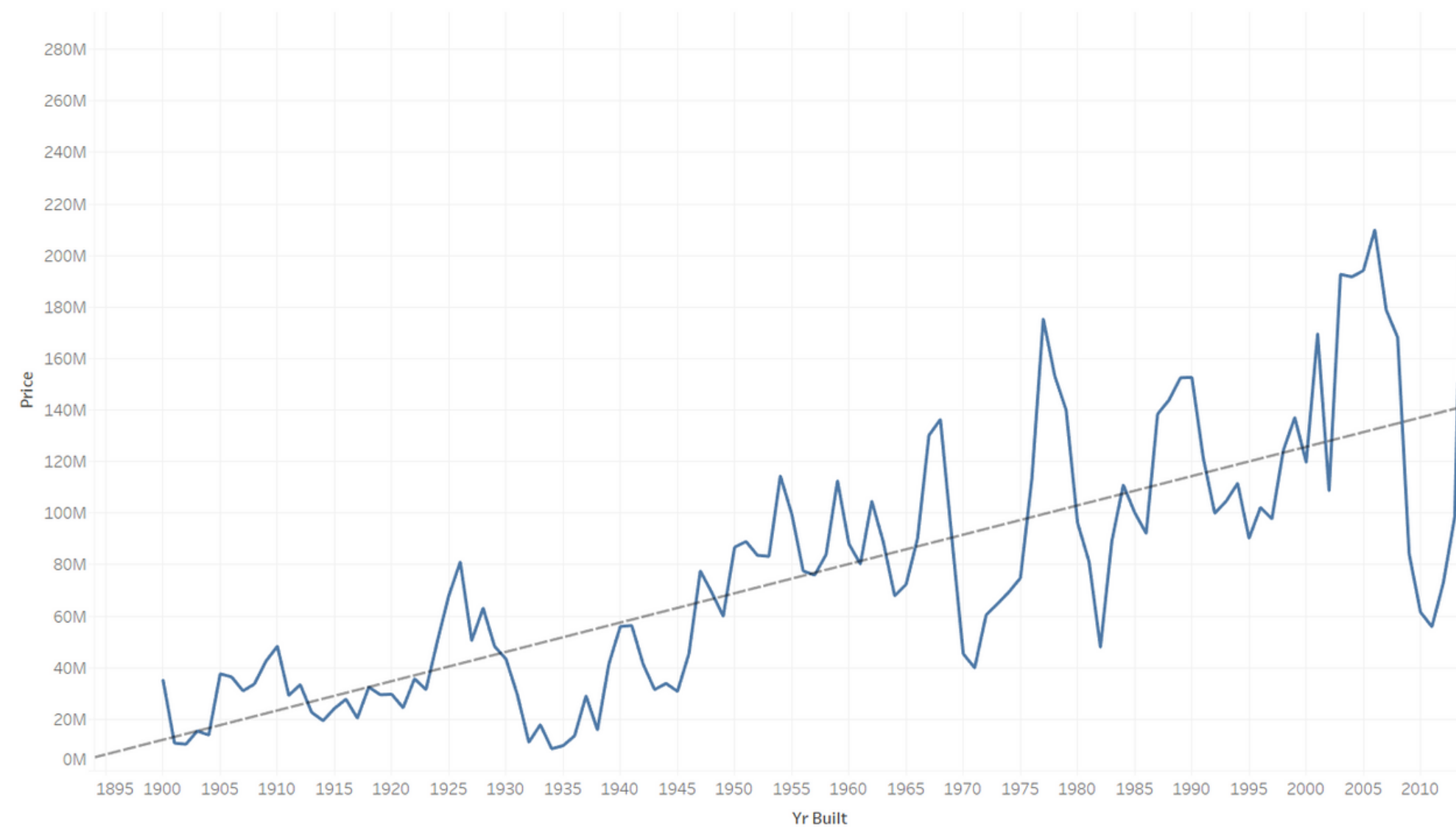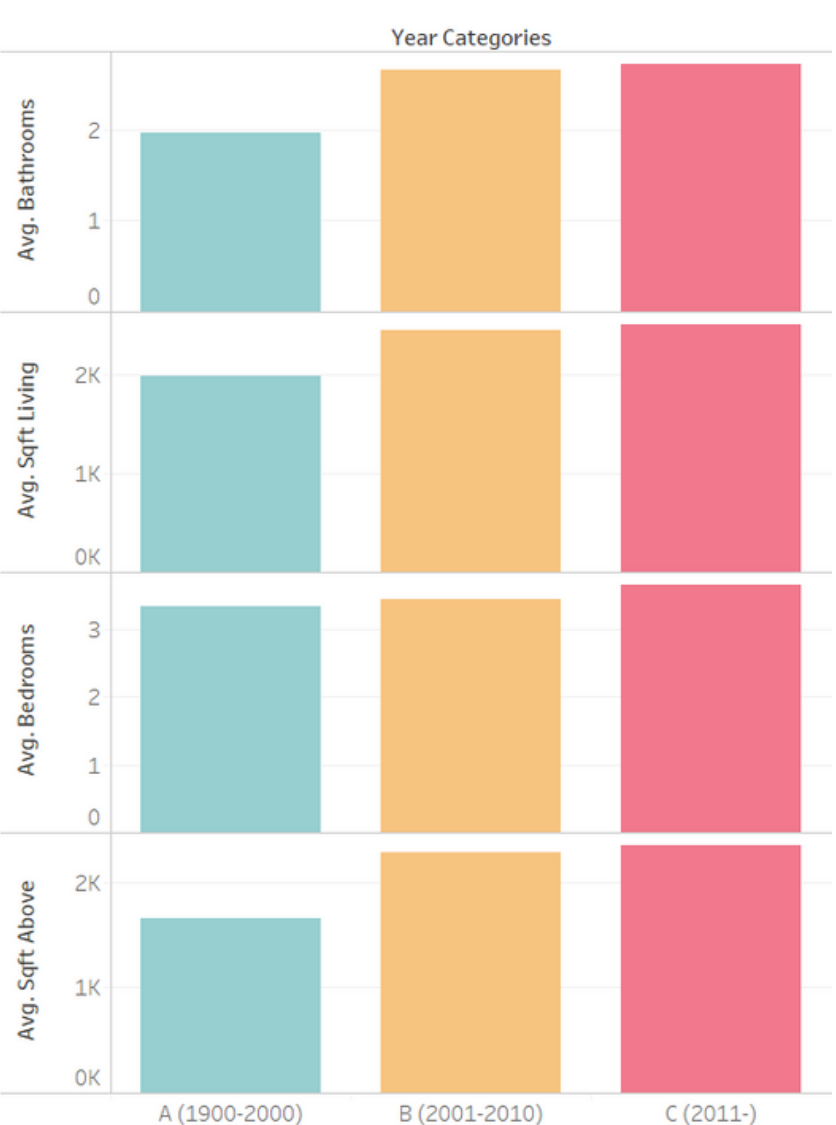
The Data

Average Values for Area:

- Price: $ 540296.57
- Year built: 1971
- Bedrooms: 3.5
- Bathrooms: 2
- Living area: 2080 ft²
- Lot area: 15099 ft²
- Grade: 7.7

The Visual Analysis
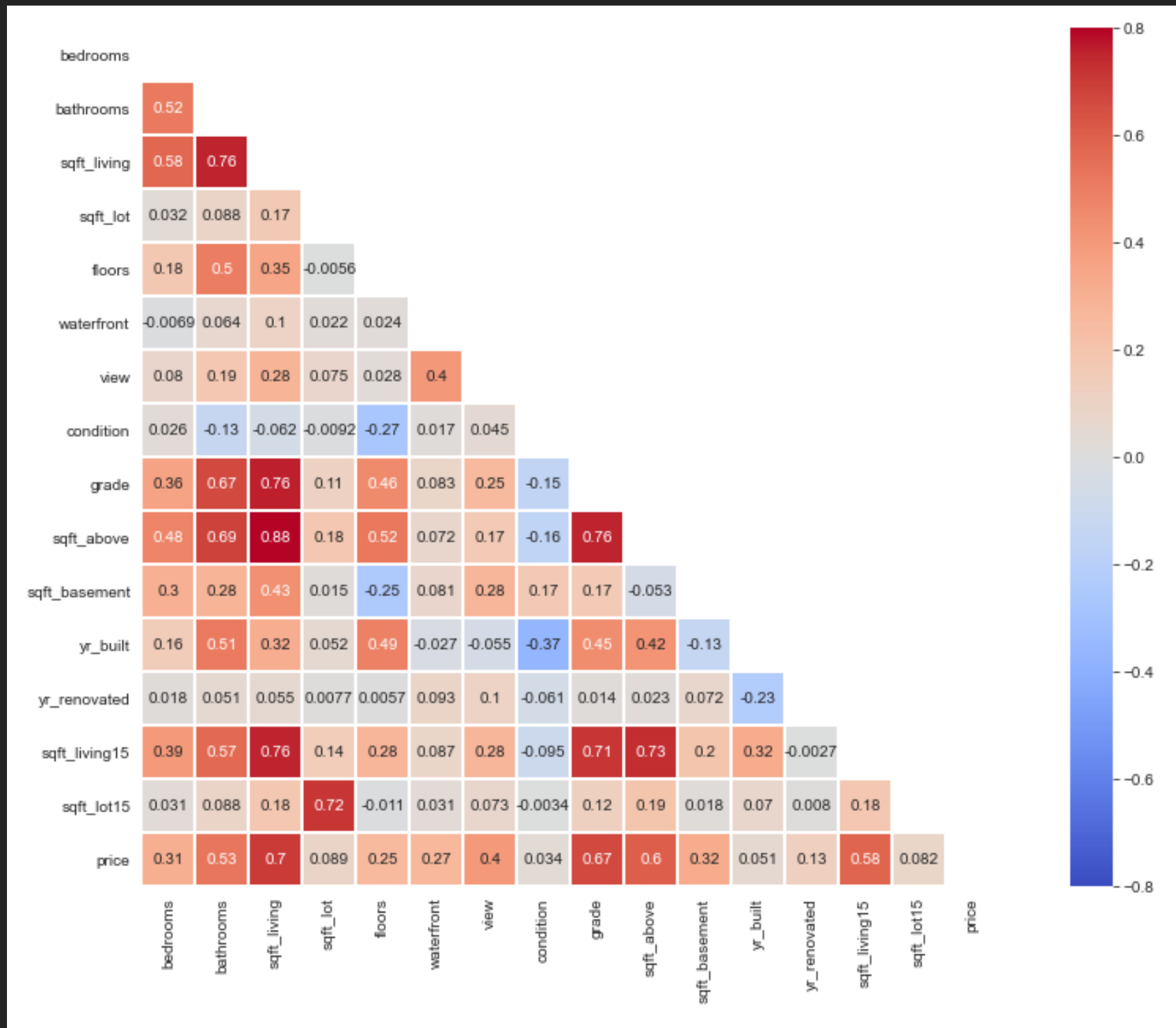

Year Bulit vs. Price (3-4 BDR)

Trends in different times

Upper Floor ft² vs. Price

The Model (EDA)

Dealing with duplicates

Correlations Matrix / Plots

Feature choice

Conversions

Outliers / Scaling

The Model (Findings)

LinearRegressor

```
lm = linear_model.LinearRegression()
model = lm.fit(X_train, y_train)
predictions = lm.predict(X_test)

exploratory run: Score: 0.6494360454513469 (all data)

second run: Score: 0.5982473698296966 (only numeric data)

after cleaning and standardizing:

Score: 0.6038192277715362 (test_size=0.2)

Score: 0.604546687253582 (test_size=0.4)
```

KNN

```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
classifier.fit(X_train, y_train)

KNeighborsClassifier()

score 1 (nn 5) : 0.003968253968253968
score 2 (nn 15) : 0.004901960784313725
score 3 (nn 55) : 0.00665266106442577
```

RandomForest

```
rf = RandomForestRegressor()
model = rf.fit(X_train, y_train)
predictions = rf.predict(X_test)

Score: 0.6356059689432338
```

The Conclusion