Philadelphia's Real Estate Pricing Model

Team 66: Daniel Gunnell, Raghuveer Krishnamurthy, Mitchell Marois, Malika Redzhepova, Megan (Aubry) Snow, Shangci Wang

Summary and Introduction

We analyzed real estate property datasets for **Philadelphia** to build a **user friendly interface to predict housing prices**. Current prediction approaches rely on recently sold properties in a similar neighborhood, ignoring some key drivers such as supply/demand dynamics and macroeconomic factors and are typically presented in a simple text format to prospective buyers. Our team used **analytical models** and **interactive visualizations** to **present real estate market predictions**.

Datasets Alignment

While in the public domain, it is very difficult to find detailed non-commercial datasets online, we were able to find transaction and assessment data for the **City of Philadelphia** via their open data API initiative. This dataset was complemented with **macroeconomic factors** such as interest rates and inflation to build our model. Overall, we downloaded and processed >3 GB of data across ~3-4 million rows of data.

Data Preparation

Initial data cleaning was focused on **removing outliers and data entry errors.** Datasets were merged to **perform explanatory analysis** and **understand the relationship of various factors.** This step was revisited iteratively as we built the models and completed the analysis.

Observations and Analysis

Philadelphia dataset was explored using **Multi-Regression, Random Forest, Gradient Boosting and Support Vector Machine** models.

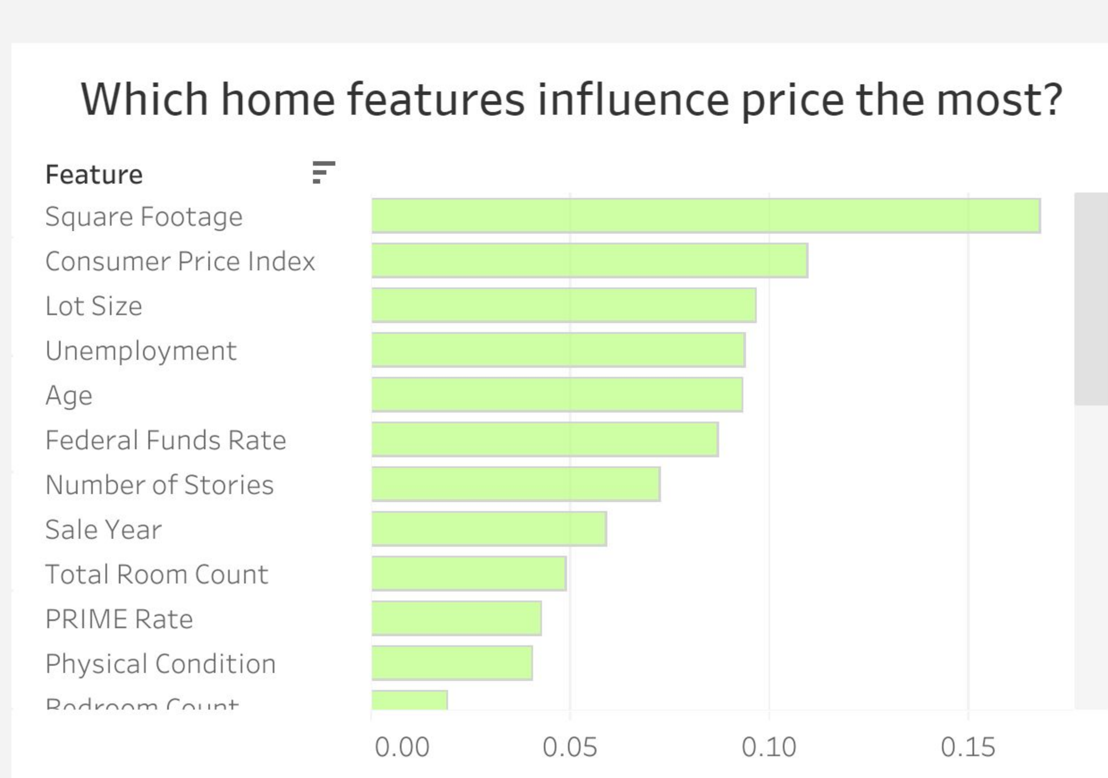
Model Performance Evaluation

Model	R^2	MAE	RMSE
Multi-Regression	0.19	125406.1	584101.7
Random Forest	**0.74**	**54500.2**	**159480.1**
Gradient Boosting	0.65	69507.0	169781.0
Support Vector Machine	0.57	94301.0	192915.0

Random Forest performed better than the other models on this dataset. Thus, in our visualizations, we will **use the Random Forest model outputs as the predicted outputs**.The training time for Random Forest was 5.1 min due to the large size of dataset.

R^2 and RMSE of Random Forest were comparable or better compared to previously published results (R2: 0.64-0.9, RMSE: 449111.5 to 1435810.8) using **hedonic price and neural network models** (Limsombunchai, Gan, Lee 2004).

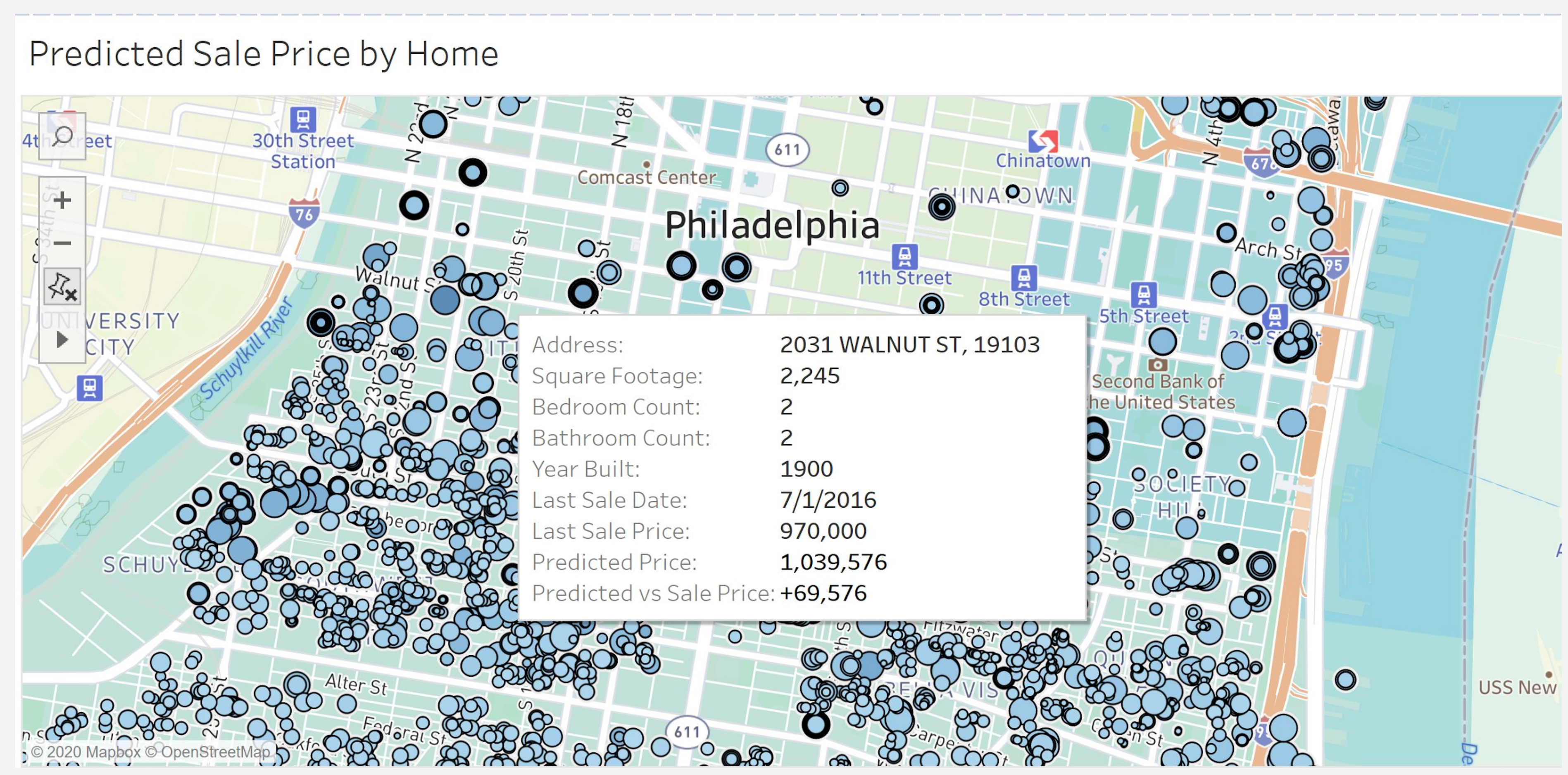
Unsurprisingly, **Square Footage influences positively prices the most.** However, economic factors such as CPI and interest rates also affect real estate values. In the case of **unemployment and interest rates, the effect is negative.** Our work is more useful than commercially available solutions today because it explicitly describes what influences prices.



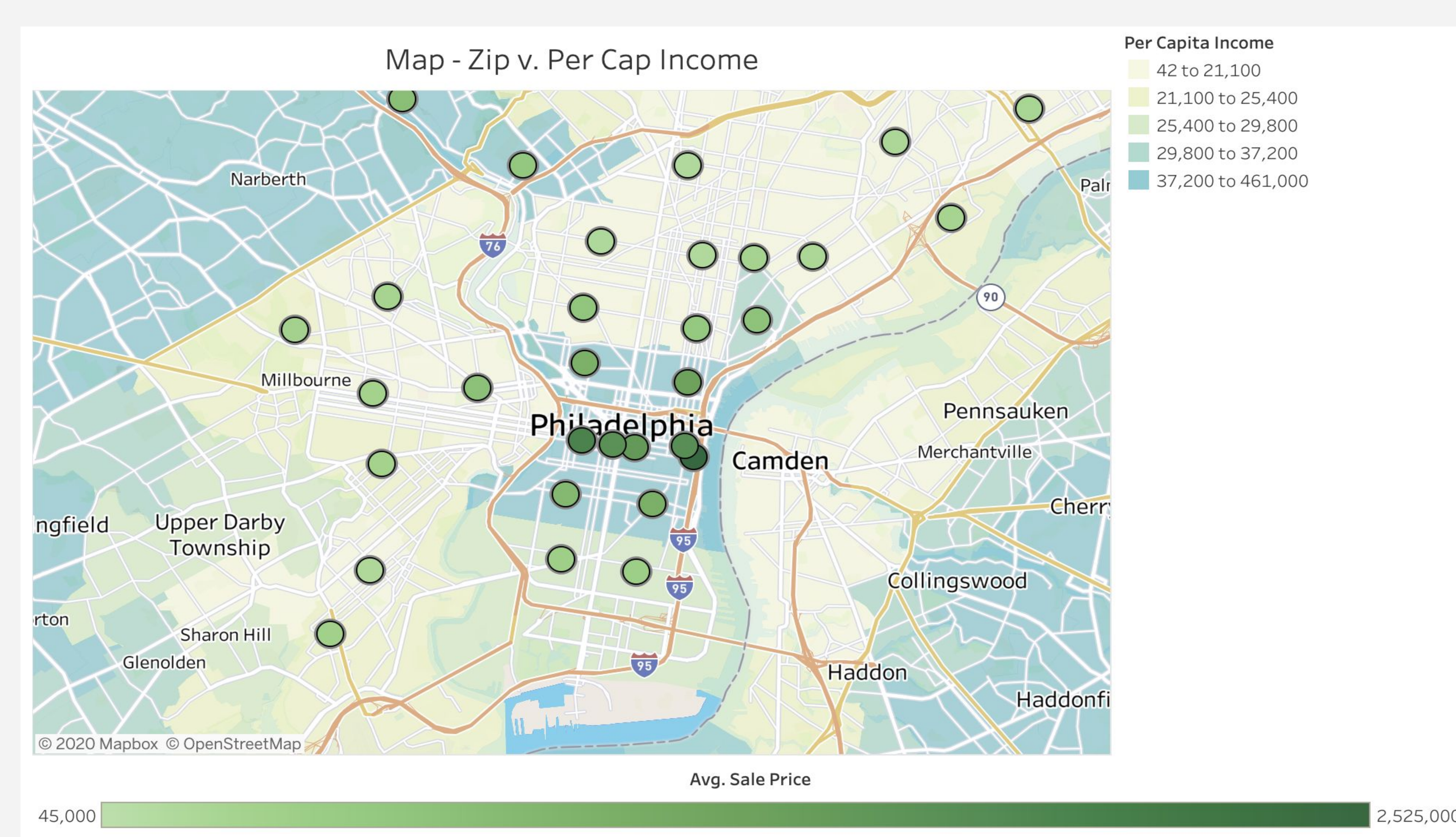
Our Approach

Visualizing Philadelphia's Real Estate Dataset

We leveraged **Tableau to visualize the results of our analysis and to create a user interface** for others to assess and understand the underlying data. One such visual is shown below in which the user can see all homes sold in the dataset and, upon hovering over an individual home, can view various metrics such as the bedrooms, bathrooms, address, last sale price, and most importantly, the predicted sale price if it were listed today.



In Philadelphia, properties **closest to downtown attractions have higher sale prices.**



On average over 2000-2020 timeframe, **Philadelphia's housing market appears to be strong.** Area shaded in light blue is our forecast of the market in the coming few years