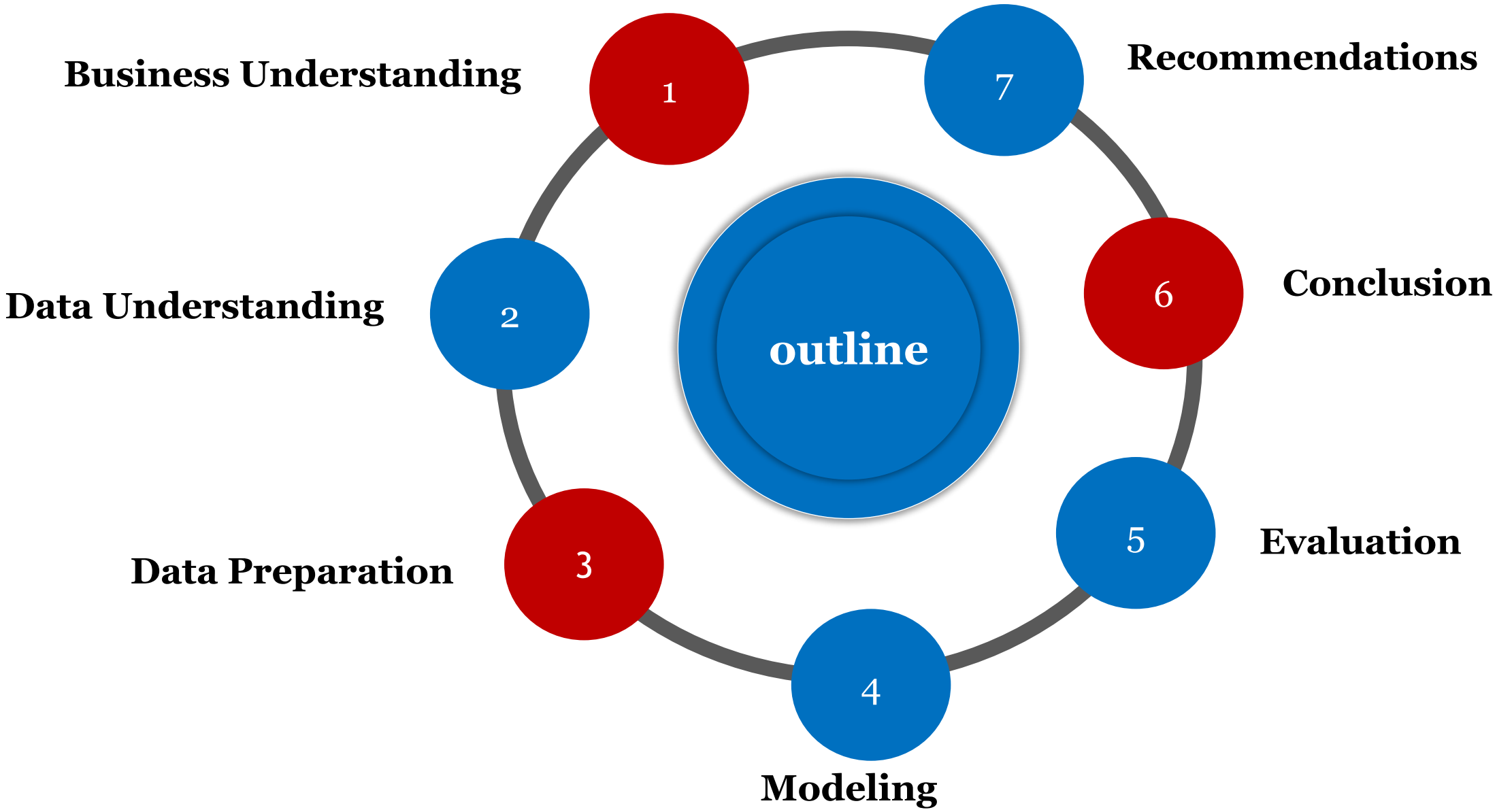
HOUSING TIME SERIES ANALYSIS

Group 8 Members

- Sammy Sifuna
- Julius Charles
- Waruchu Kuria
- Rael Ndonye
- Alan Omondi
- Janet Khainza


Business Understanding
1
Recommendations
7
Data Understanding
2
outline
6
Conclusion
Data Preparation
3
5
Evaluation
4
Modeling

Overview

- **Market Transformation:** Initial rise in home prices driven by real estate's **perceived long-term value** and a decline of home values due to the **Great Recession diminishing buyers trust.**

- **Restored Stability:** After the recession, stability was restored due to **low interest rates** as well as **urbanization** and **technology** which attracted younger dwellers to cities.

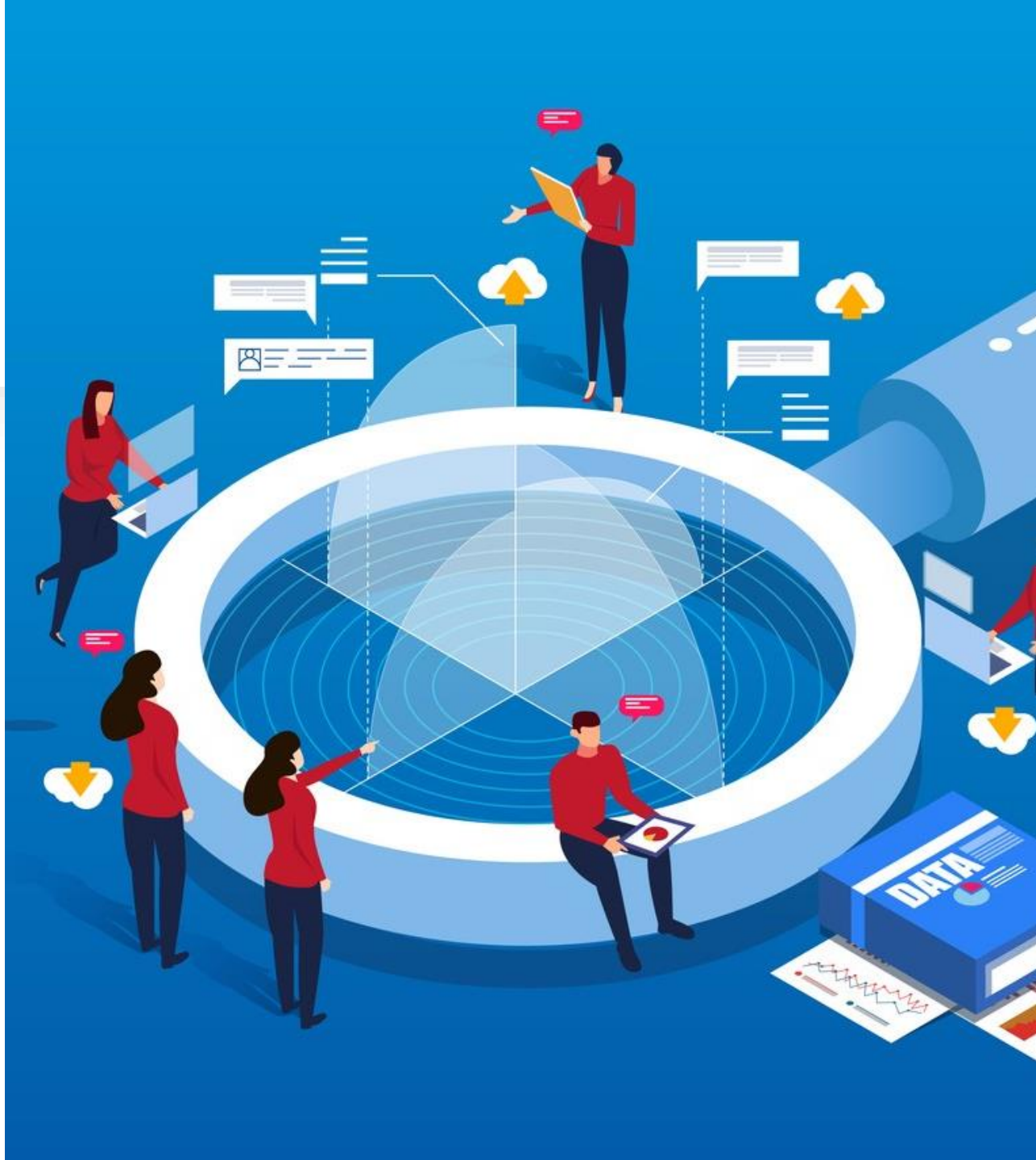
Business Problem

- Waridi financial Investment firm has a big decision to make on the **best 5 zip codes to invest on short-term basis.**
- Analyze **median monthly housing sales prices** for over **14,000 United States zip codes** and find the best return on investment with tolerable risk.
- Calculating **22-year, 5-year and 3-year ROI** to figure out the best zip codes giving the most profit.

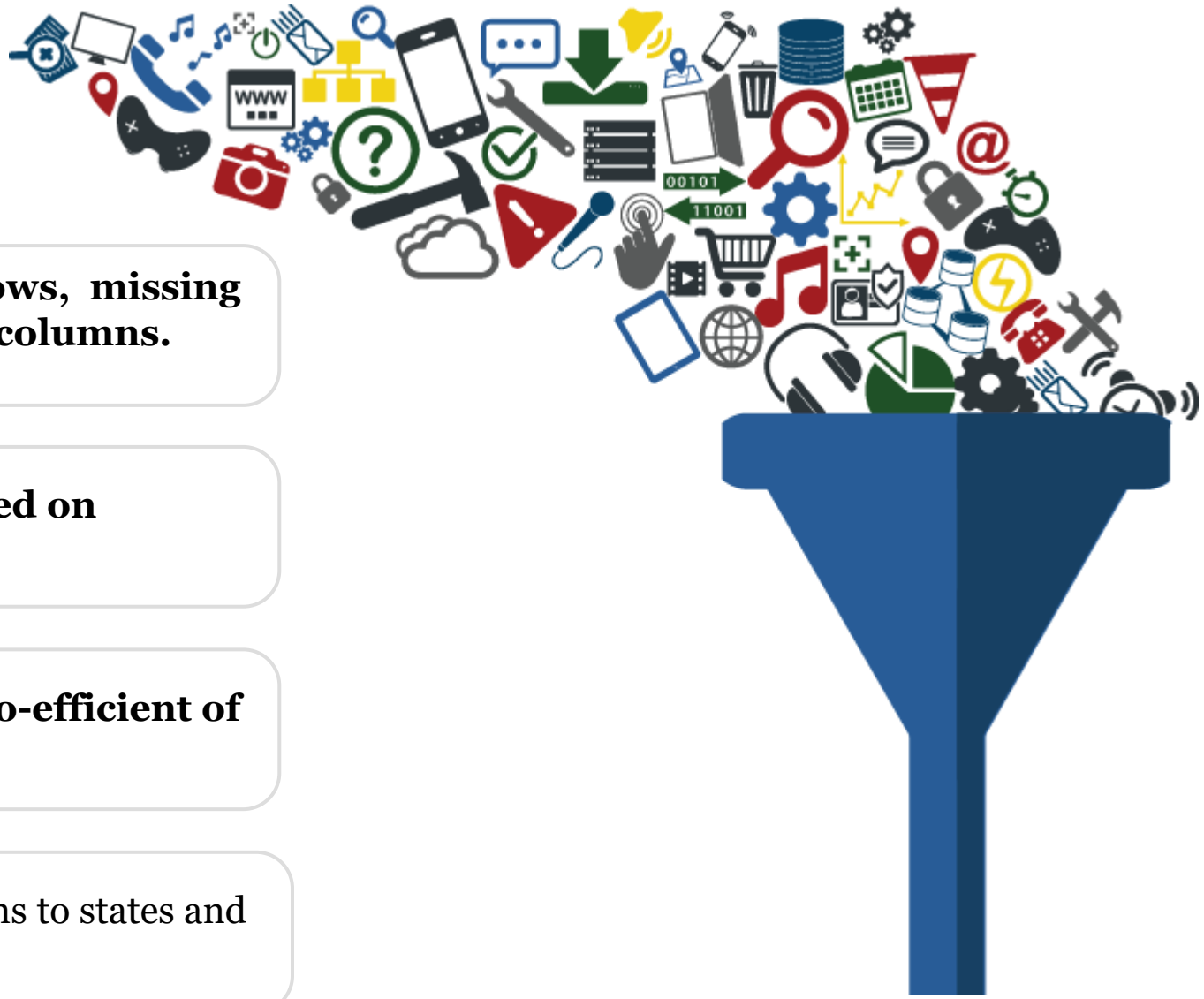
- To determine the top 5 zip codes that show the highest potential return on investment (ROI) in 5 years.
- To identify which top 5 states have the highest ROI.
- To determine the best month to sell property to maximize ROI.

Data Understanding

- Median monthly housing sales prices data for **265 zip codes** in the USA, over the period of **April 1996 through April 2018.**
- Dataset contains **14,723 records** and **272 variables**
- The features contain a mix **of numerical** and **categorical variables**
- The data will be used for time series analysis to **predict housing sales prices.**



Data Preparation

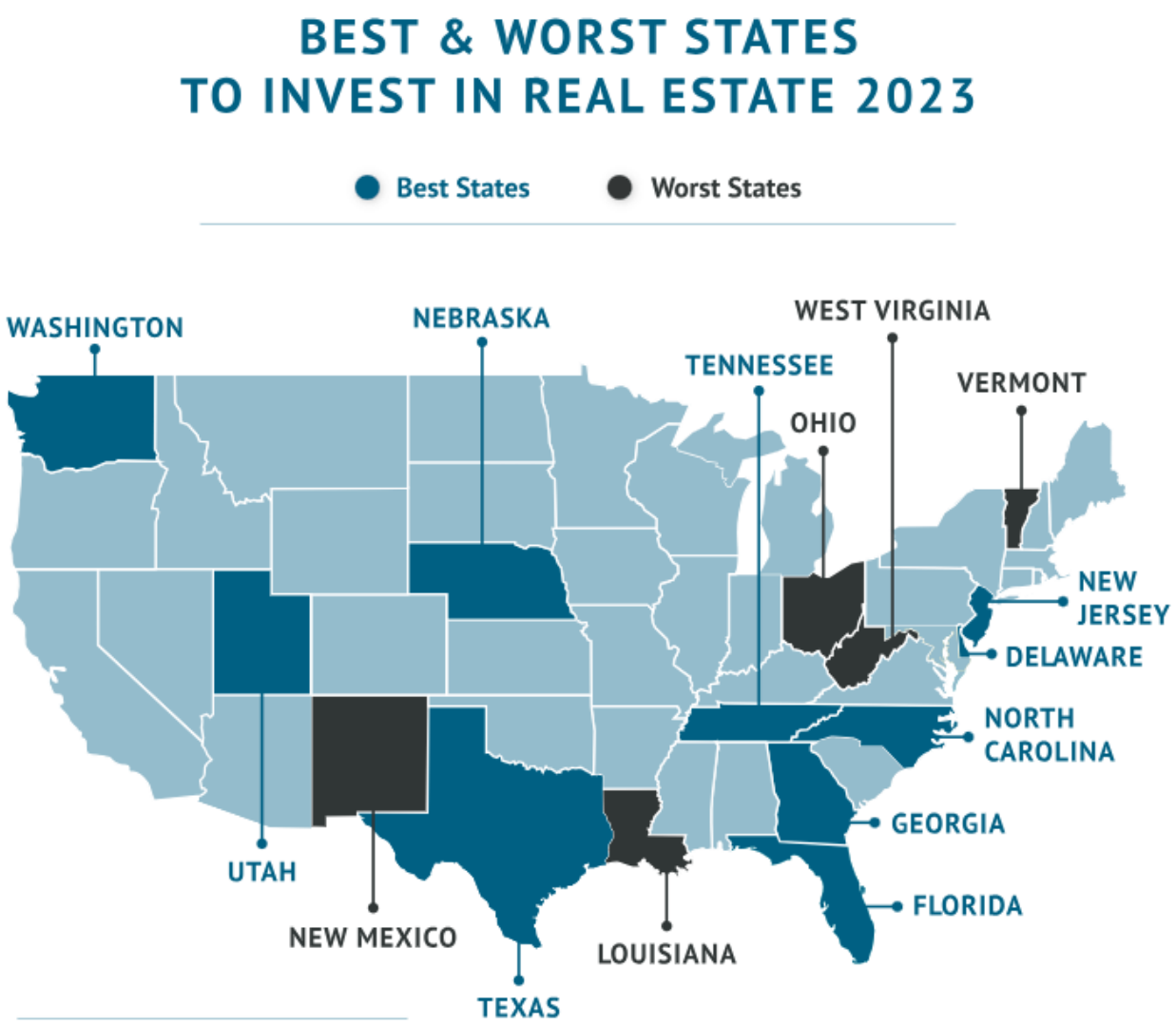


1. Data cleaning including checking for **duplicated rows, missing values**, **dropping unnecessary and renaming columns.**

2. Data analysis involved selecting **top 10 states based on domain knowledge** of the best states to invest in.

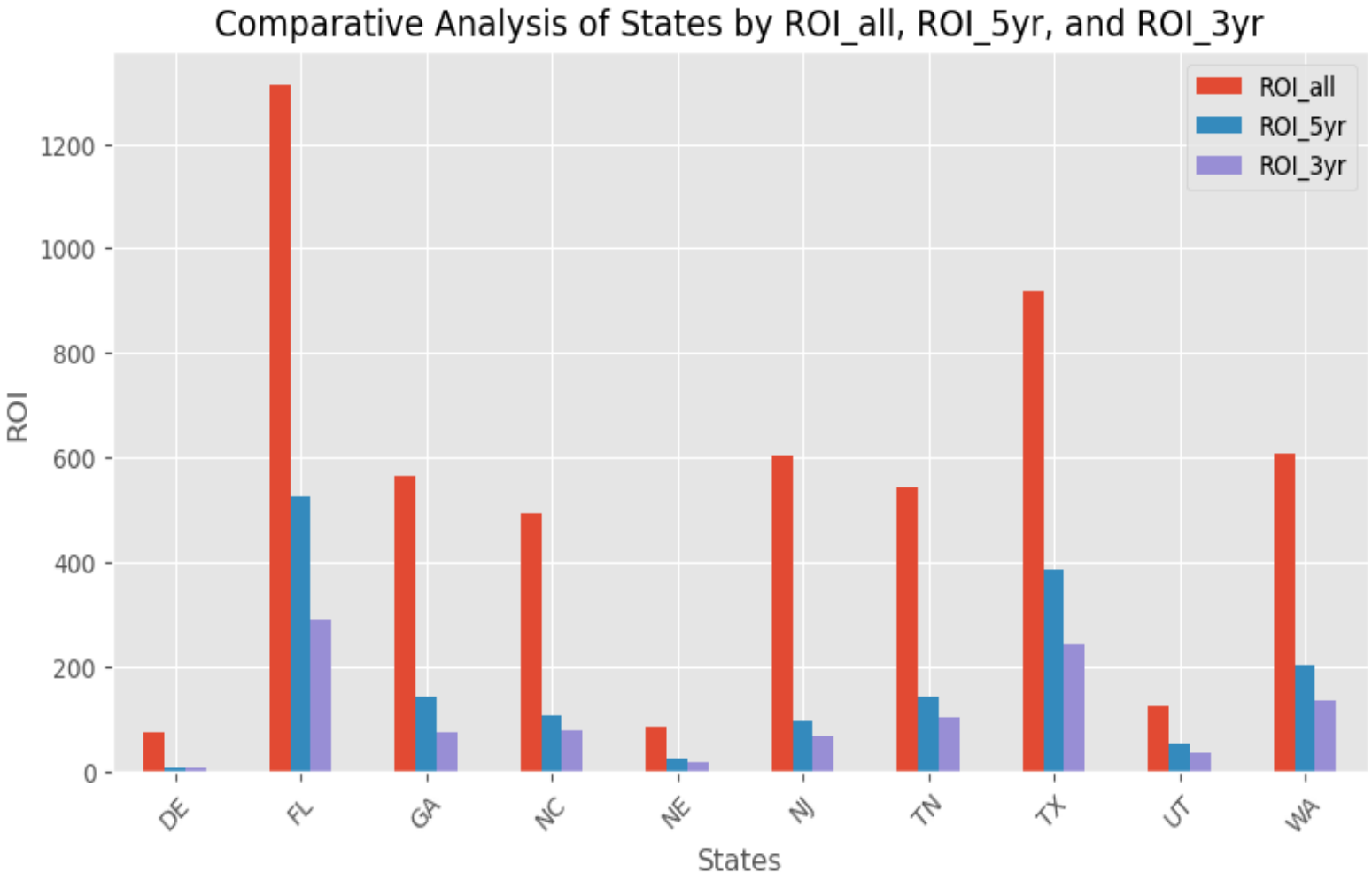
3. Calculating **Return on Investment (ROI)** and **Co-efficient of Variation(CV)** to determine where to invest.

4. **Created visualizations** of ROI and CV in relations to states and zip codes as well as time to average home sales price



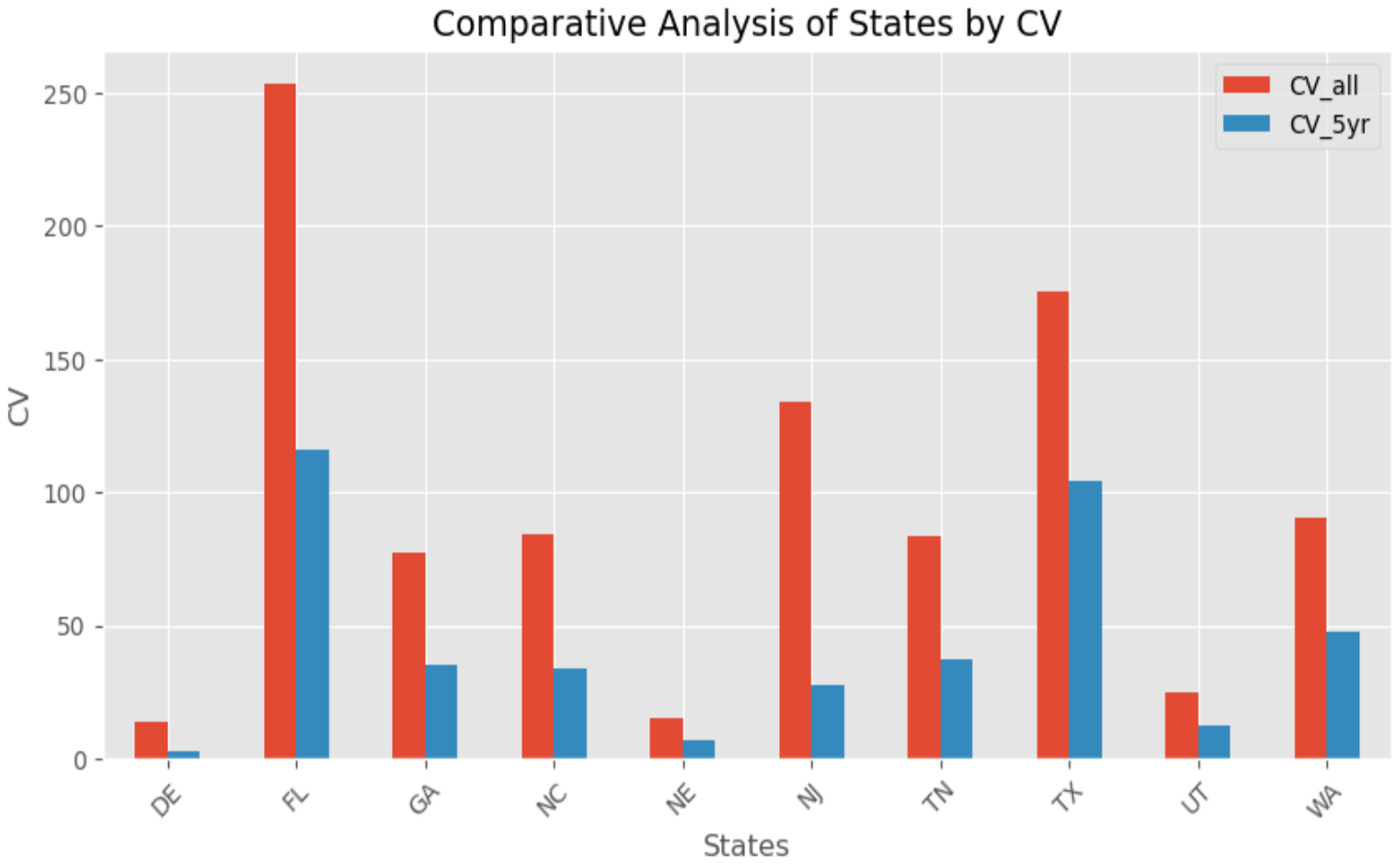
- High population of renters
- Demand for more housing
- Low cost of living
- High real estate market appreciation
- Job opportunities
- High happiness scores

States VS ROI



- **State of Florida** has the **highest ROI** for all specified periods.
- Even though New Jersey comes third in 22-year ROI, **it performs poorly in the 5-year and 3-year ROI**
- Delaware State has the lowest ROI values for all periods.

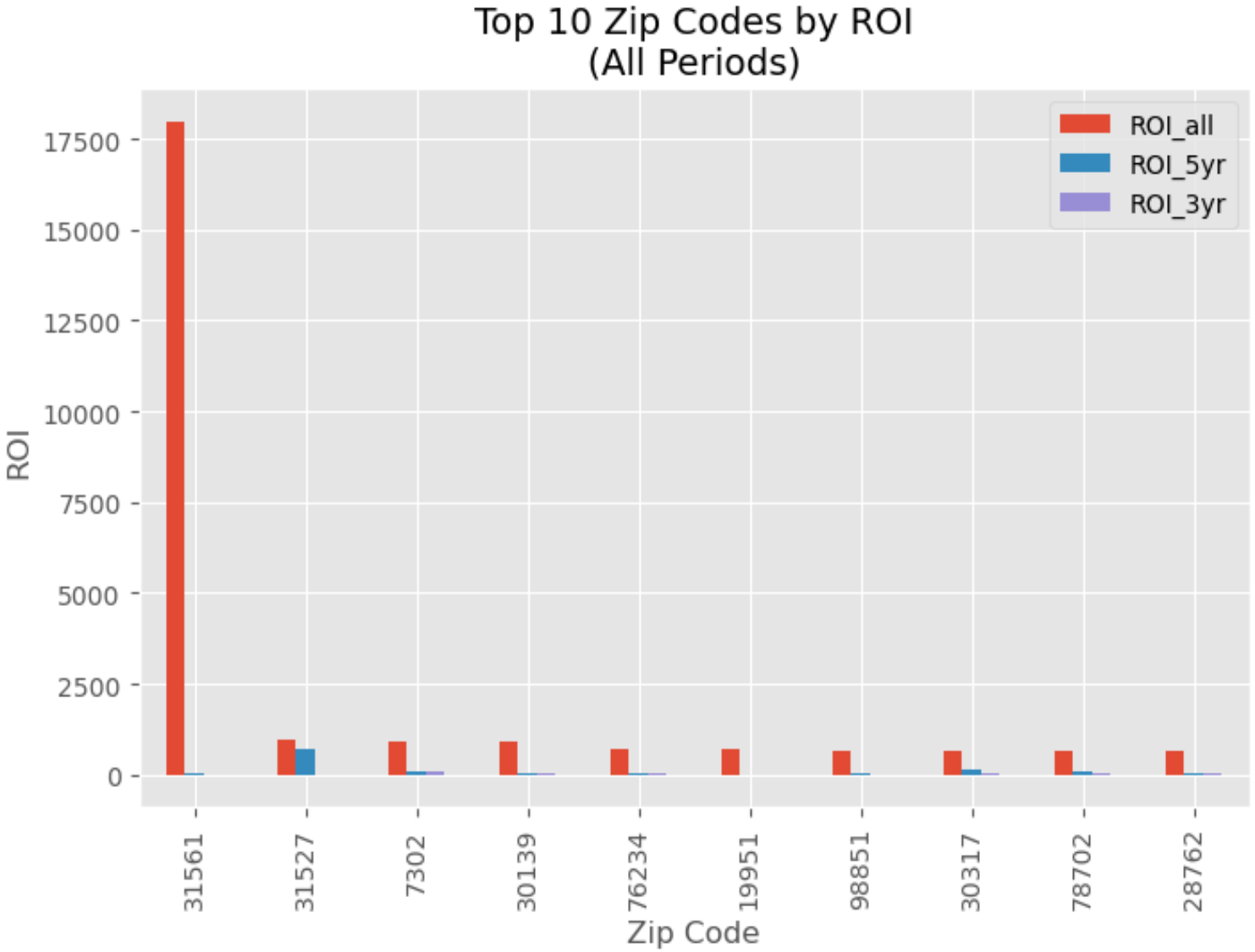
States VS CV



- Again, Florida State has **the highest Co-efficient of Variation** values meaning that though it has **high return on investment** it is also **highly risky to invest.**

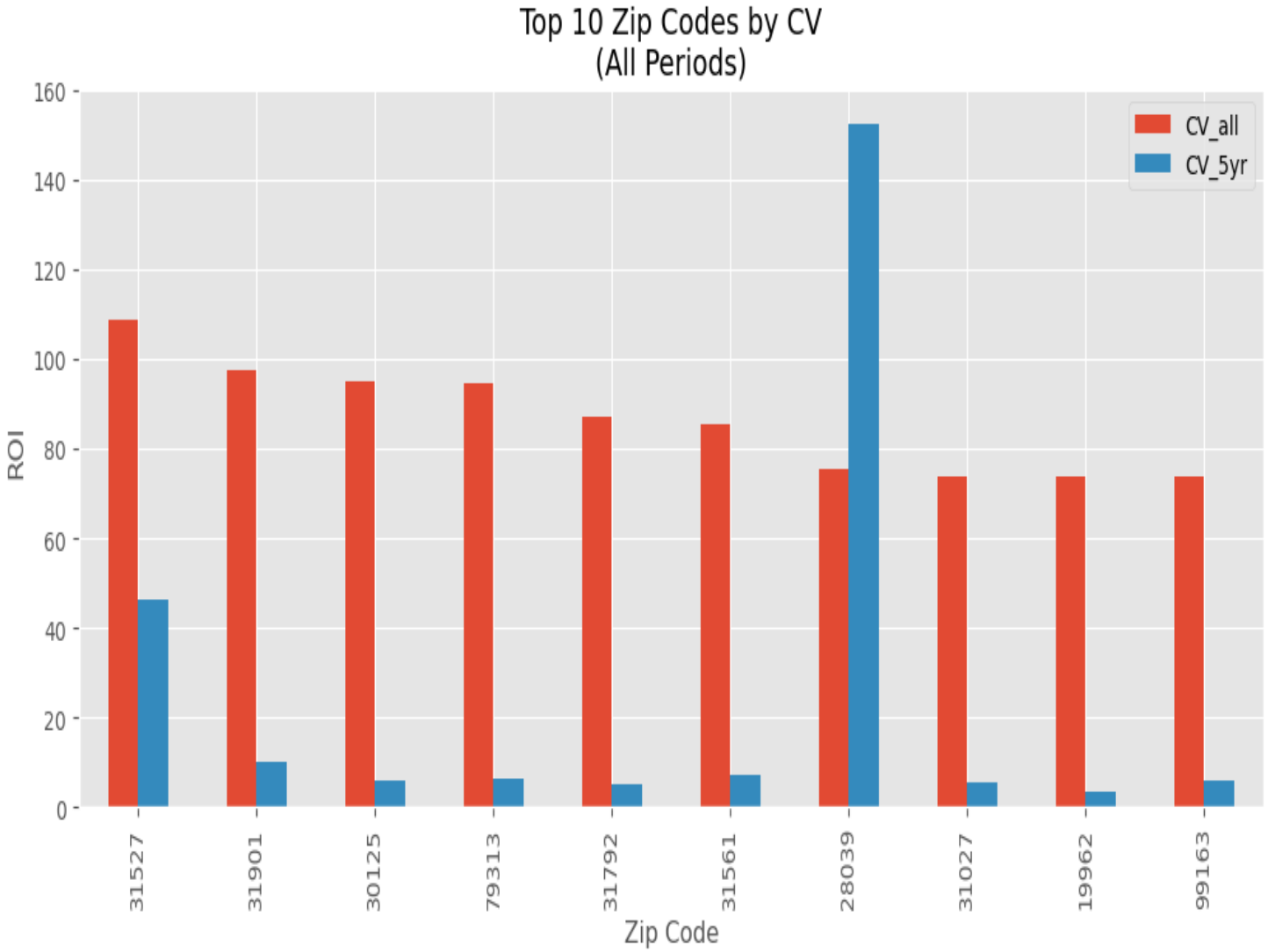
- Delaware has **low returns and low risk.**

Zip codes VS ROI (all periods)



- Zip Code **31561** (**Sea Island, Georgia)** seems to be performing extremely well in terms of potential for profits.
- Zip Code **31527 (Jekyll Island, Georgia)** has both considerably higher historic and 5-year ROI
- 3-year investment does not show so much significant ROI

Zip codes VS CV (all periods)



- Zip code 28039 (**East Spenser, North Carolina)** has the highest CV for the past five years, meaning that it **has high risk of investment.**

- **31527 (Jekyll Island, Georgia)** has the highest historical overall CV.

- The following have the least CV in the 5-year period.

i) 19962 (**Magnolia, Delaware**)

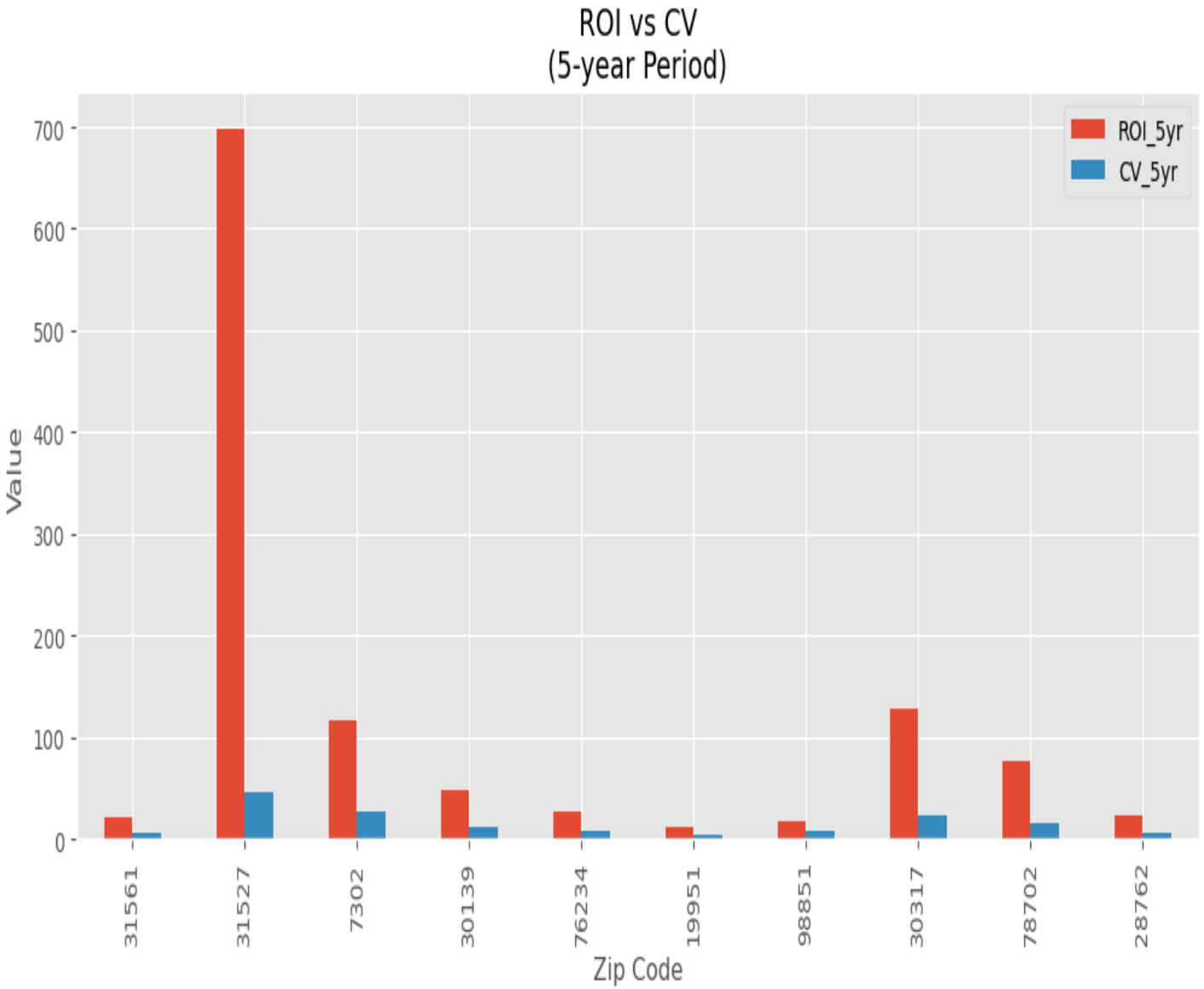
ii) 31792 (**Thomasville, Georgia**)

iii) 30125 (**Cedartown, Georgia**)

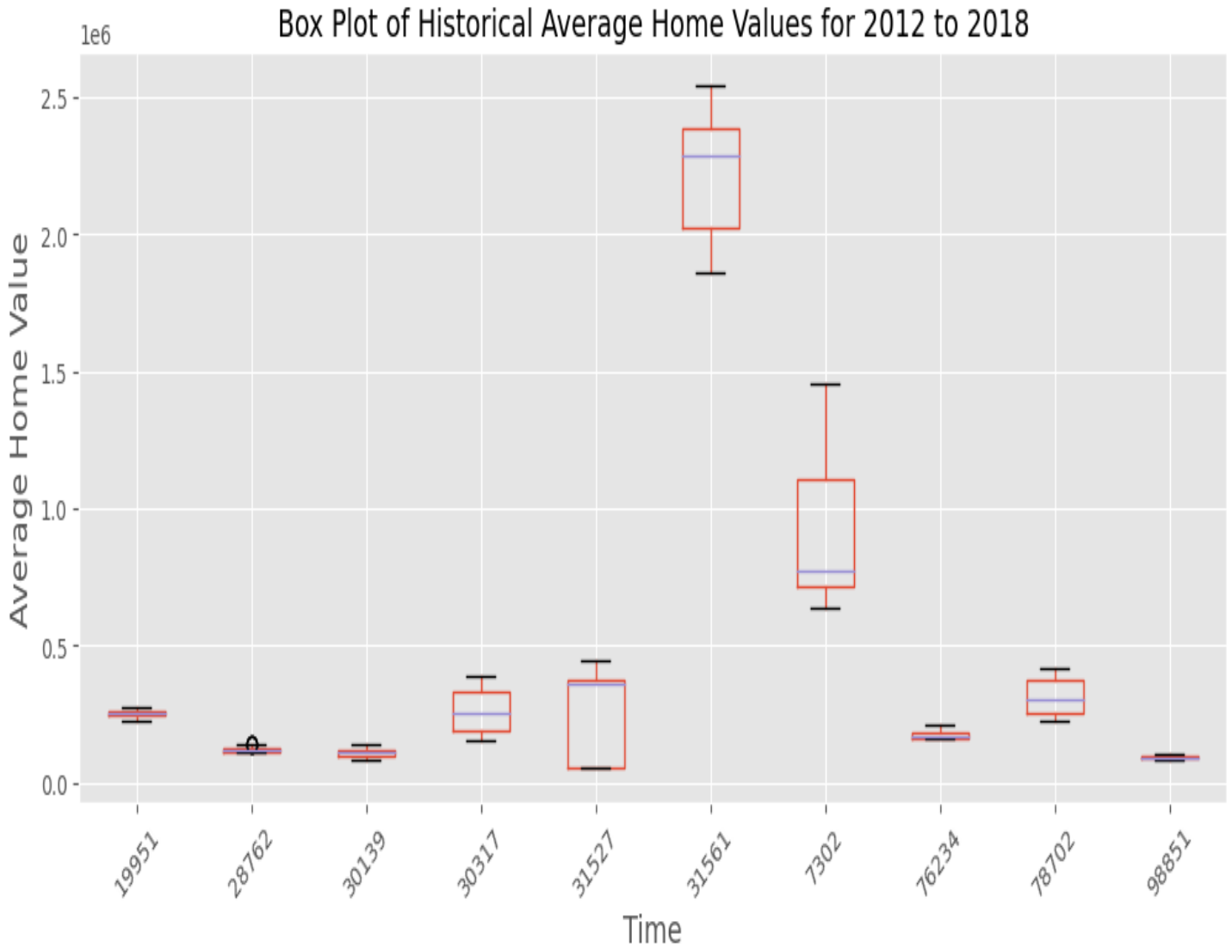
iv) 99163 (**Pullman, Washington**)

v) 79313 (**Anton, Texas**)

5-year (ROI and CV)

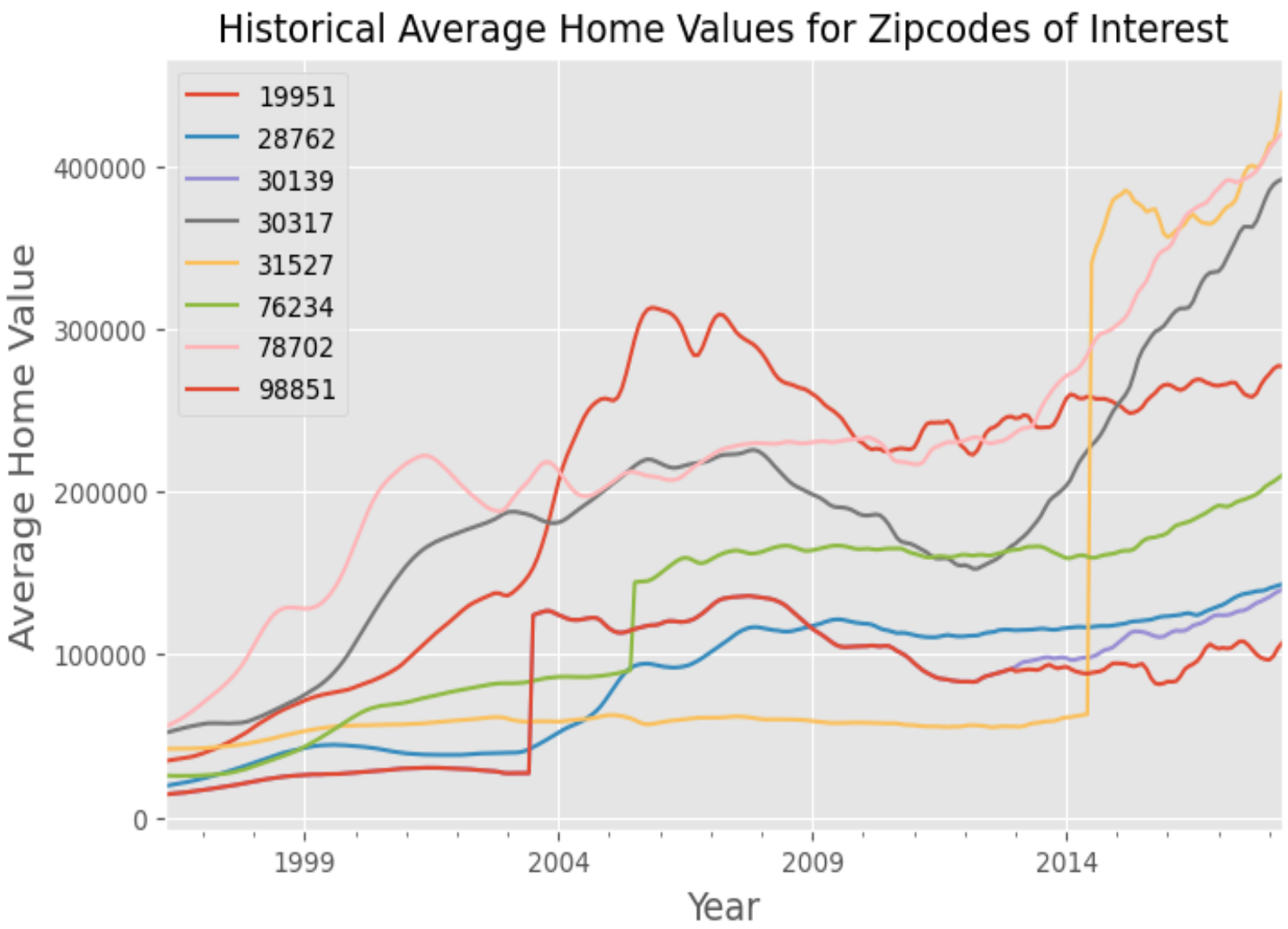


- **31527** (**(Jekyll Island, Georgia)** stands out again with a **high 5-year ROI and CV**.
- **7302** (**Jersey City, New Jersey**), **30317** (**Atlanta, Georgia)** and **78702** (**Austin Texas**) have high ROIs but higher than most CV values.
- The visualization gives a good comparable visualization and was used to form the time series for modelling.



Average Home Value per Zip code

- **31561 (Sea Island, Georgia) and 7302 (Jersey City, New Jersey)** have the highest average home values.
- The two seem to be outliers.



Average housing price vs Time

- Time series analysis shows an **upward trend from 2012 onwards.**
- There is a **general decline** in the average housing sales price **from 2008 to 2012**.

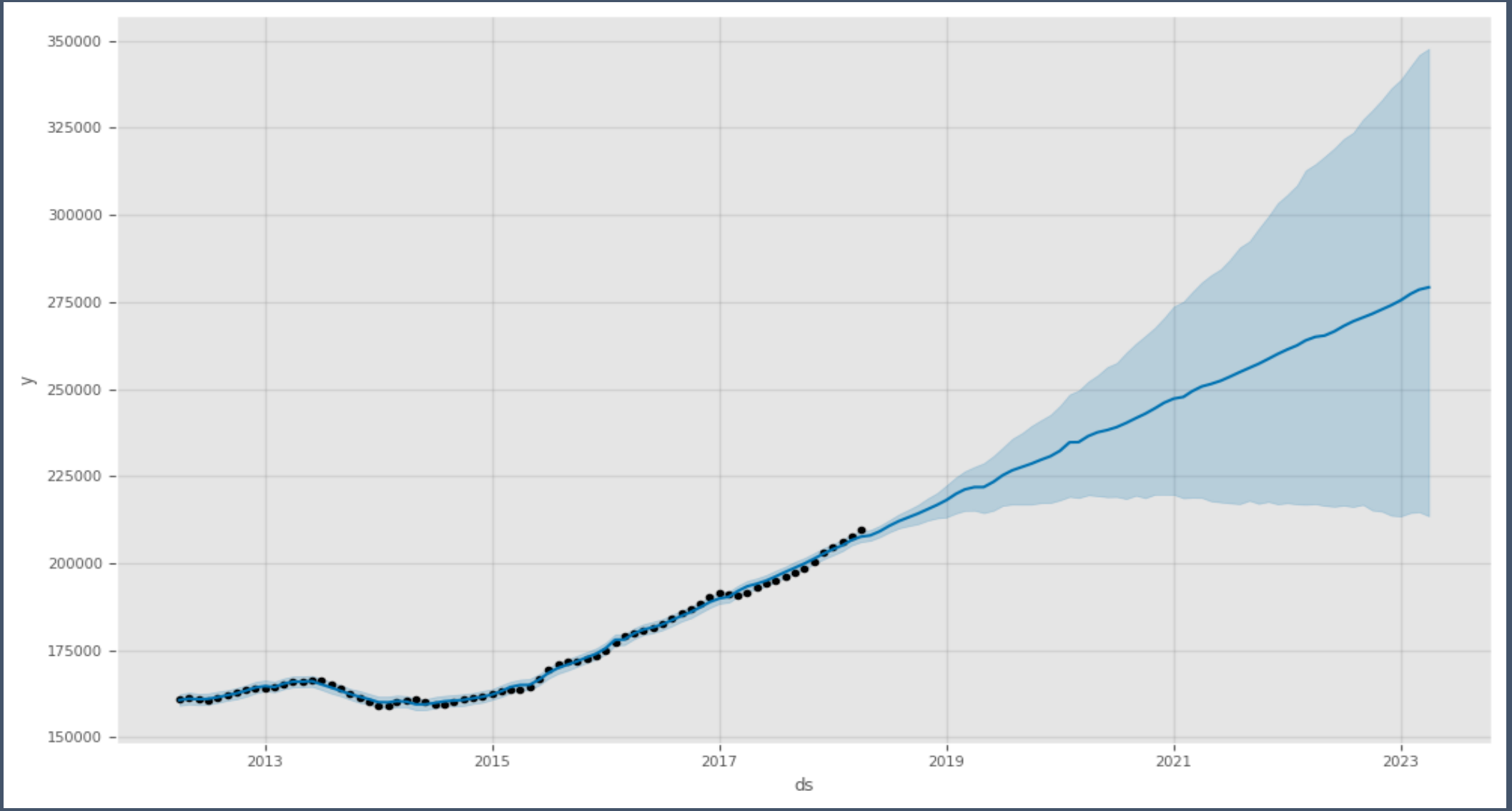
TIME SERIES MODELING

- Model used **past home sale values to predict** future sales.
- Sampled the top 10 zip codes based on **ROI** and **CV**
- Revealed the top 5 zip codes based on **RMSE and AIC metrics**
- Used **SARIMAX** and **Facebook Prophet** models

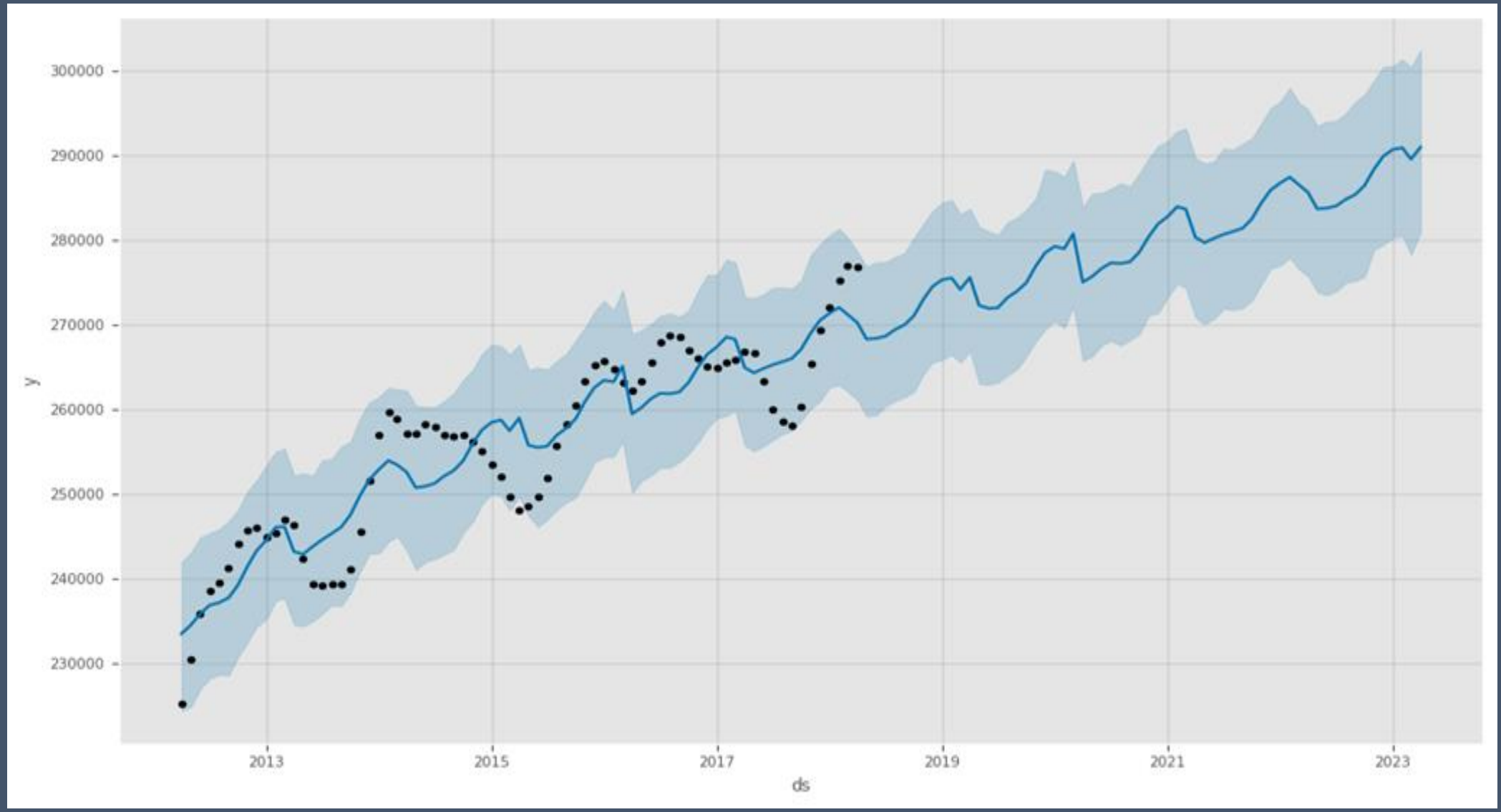
PREDICTIONS AND FORECASTING

- While both models performed the same predictions and metric calculations, Facebook Prophet model was able to forecast the time series well.
- The top five zip codes are visualized in the next slides.

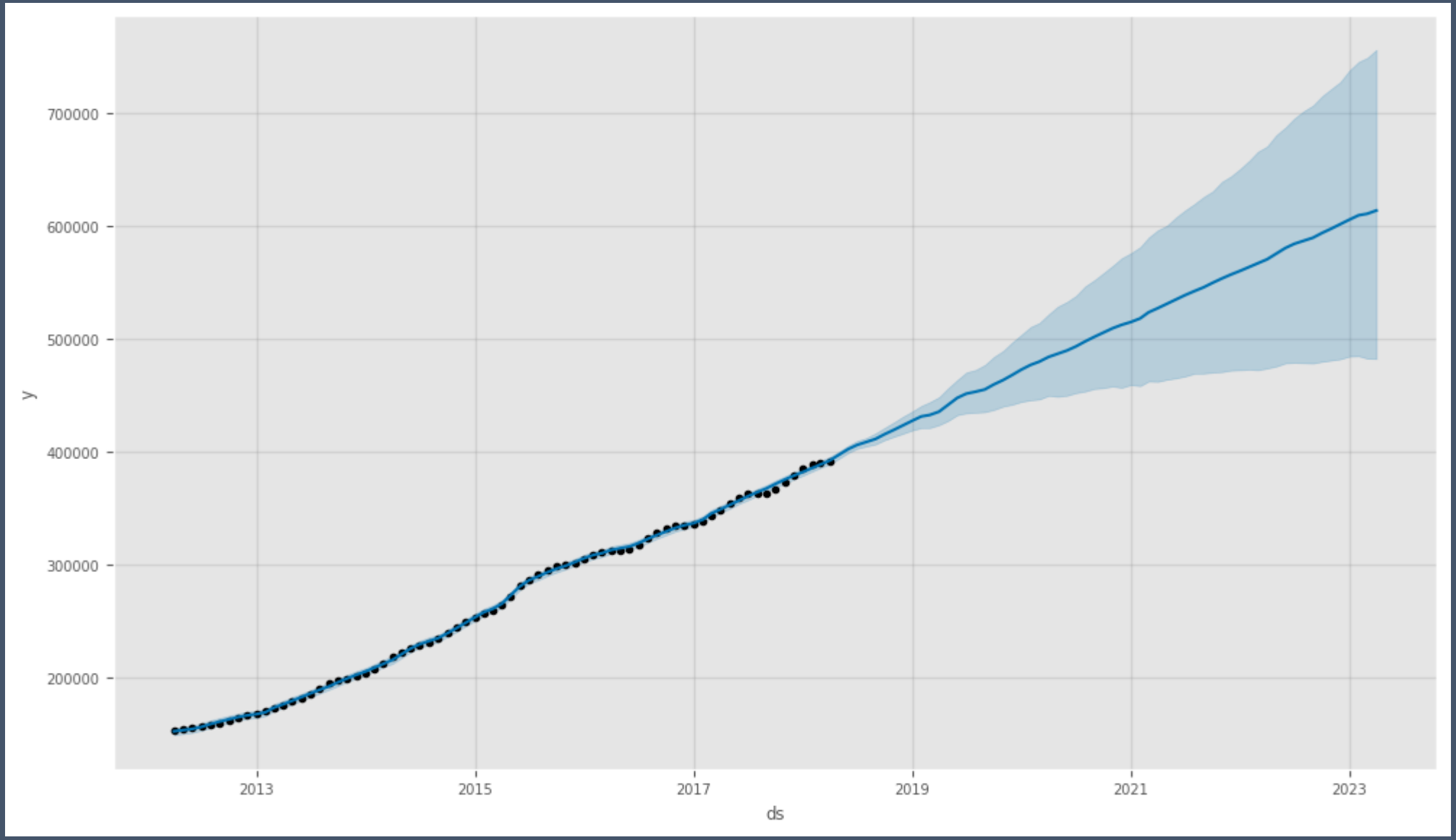
1. Decatur, Texas



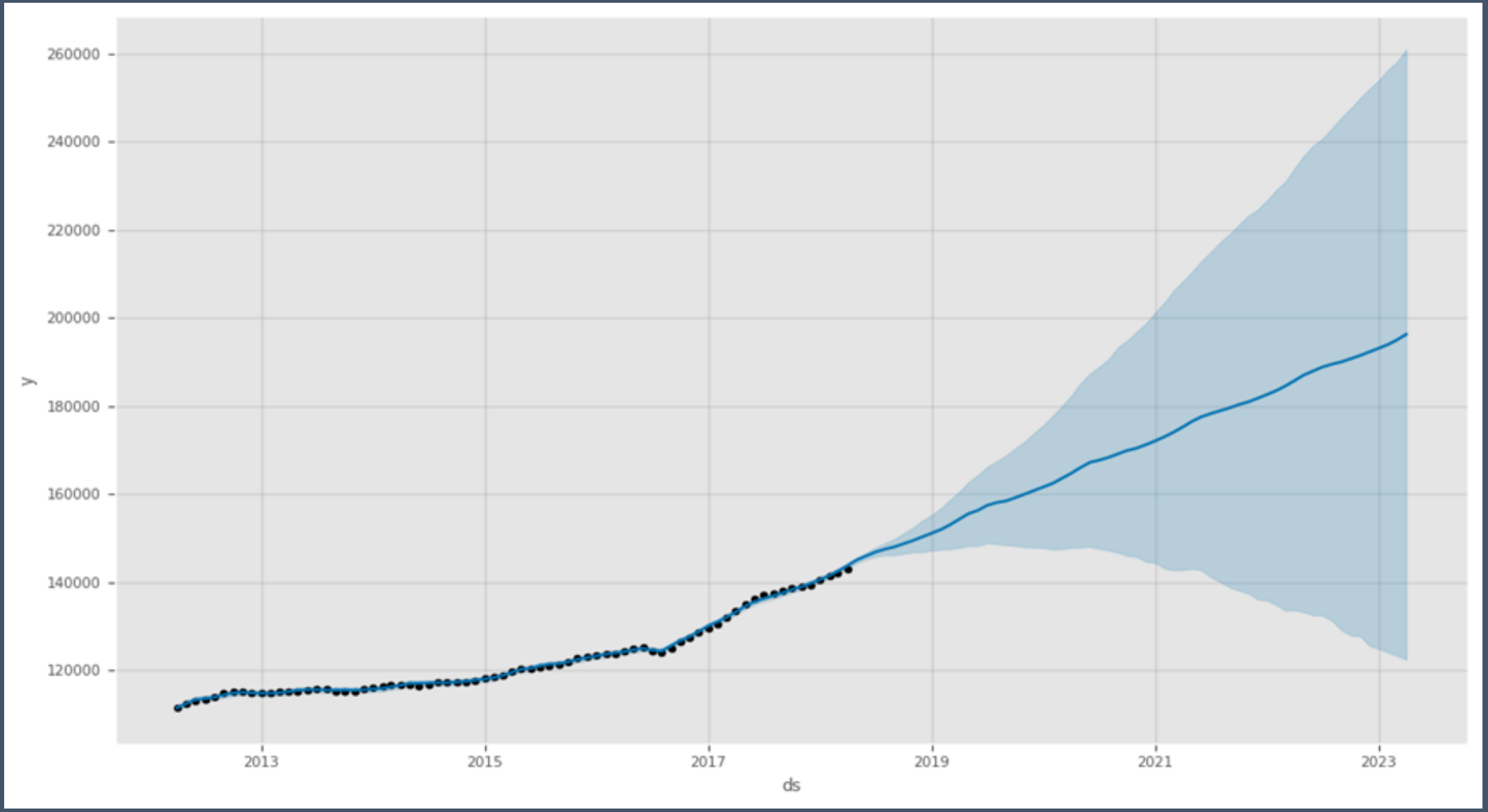
2. Harbeson, Delaware



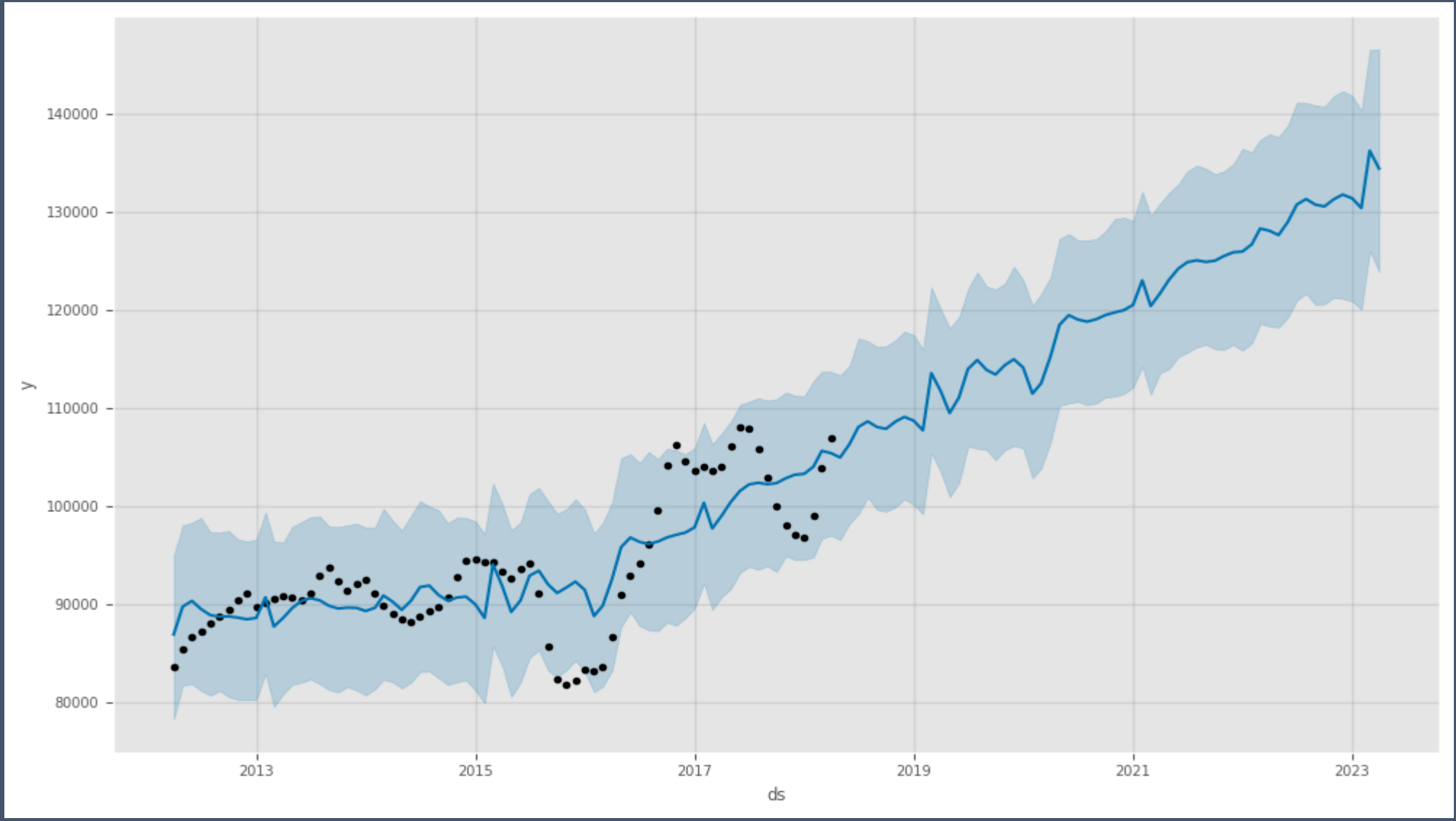
3. Atlanta, Georgia



4. Old Fort, North Carolina



5. Soap Lake, Washington



Model Evaluation

	RMSE	AIC
Decatur, TX	5033.433673	765.971035
Harbeson, DE	5753.477978	912.078435
Atlanta, GA	6154.167274	892.355686
Old Fort, NC	6779.883346	692.602794
Soap Lake, WA	10254.866297	875.084354
Austin, TX	11318.298431	1006.814736
Fairmount, GA	13161.083951	835.732403
Jekyll Island, GA	30178.717667	1201.778543

Conclusion

- **76234(Decatur, TX), 19951 (Harbeson, DE), 30317 (Atlanta, GA**), **28762 (Old Fort, NC**) and **98851 (Soap Lake, WA)** are top 5 zip codes to invest in because of high ROI and recourses should be put in them.
- The best months to sell to maximize profit are:
    i. December in **Harbeson**, DE,
    ii. April in **Old Fort**, NC and **Fairmount**, GA
    iii. June in **Decatur**, TX, **Soap Lake**, WA
- **Facebook prophet** is a better model in predicting future housing sales values.
- The return on investment and the risk are not necessarily correlated. **Sometimes a high ROI could also mean a high risk and vice versa**

RECOMMENDATIONS

- Invest on the top 5 zip codes and **allocate resources towards them.**
- Obtain **current data** after 2018 for current predictions
- There is need to **consider outside factors** (housing grade, property size, renovations) that may also affect Sales value.

Thank You!