USBusinessReport

Simone Trindade Steel

20/06/2020

US Business Report: an analysis of economic sentiment and new housing construction

This reports is based on a study of business and industry trends in the United States using the US Census Bureau data. The inspiration for this study is the hypothesis that US National level of value generation, as an indication of economic optimism, drives the antecipation of new property demand.

The analysis looks for regional variations in the US and attempts to predict the regional growth of new home construction and sales.

The source of data is https://www.kaggle.com/census/business-and-industry-reports

Understanding the data

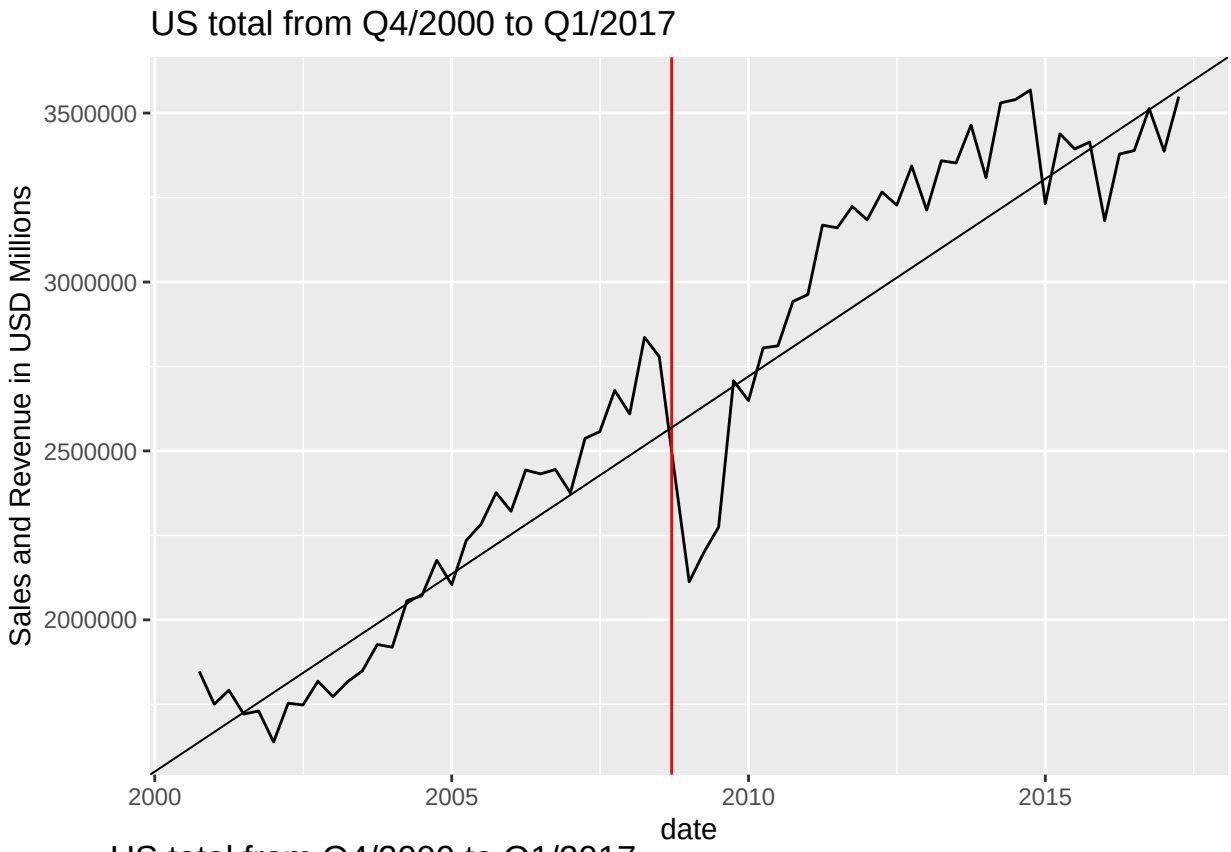
The original dataset has many economic factors that will not be used in this study. The time series that are relevant to the subject of this report are the macroeconomic indicators (Financial Reports) and the new housing indicators (New Home Sales, New Residential Construction)

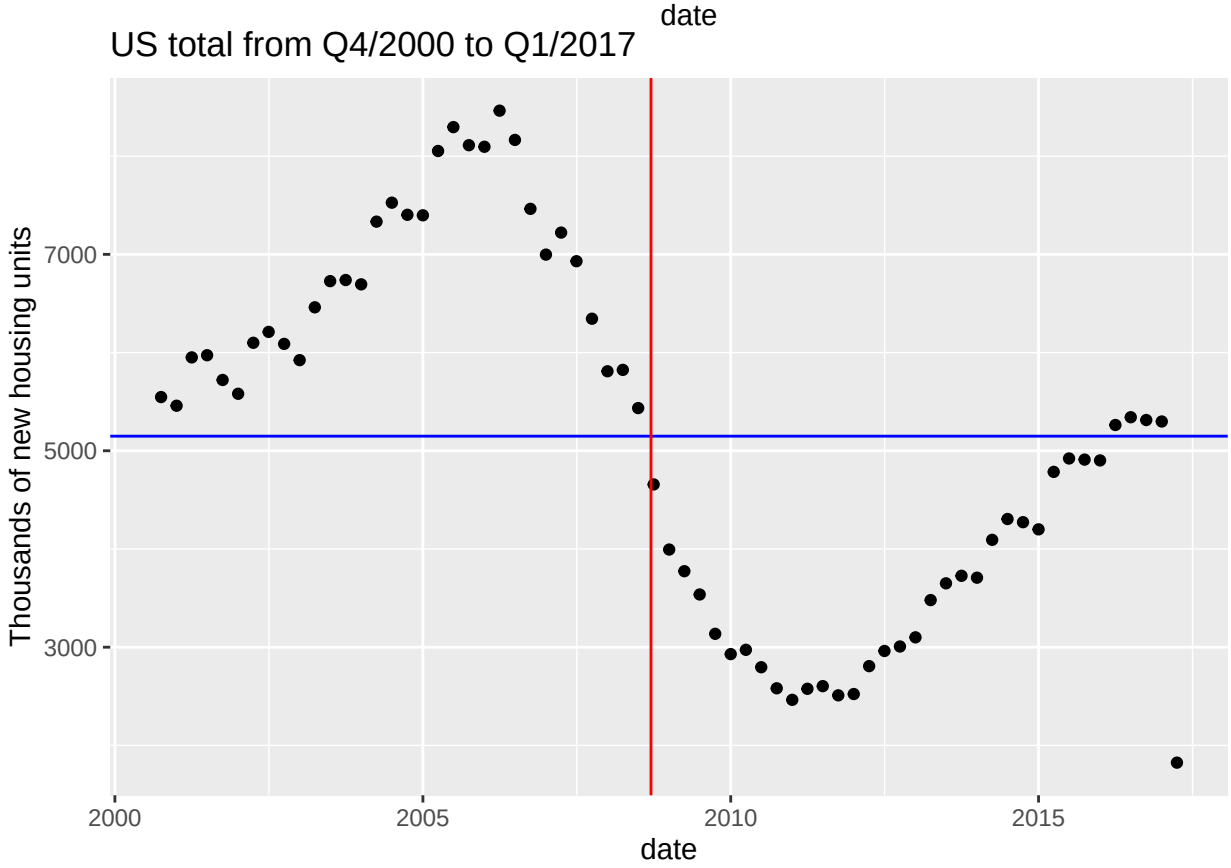
All dates indicate the begining of the analysis period, and they have been used to align the time series into quarterly periods.

The graphs below summarise the US national totals, and the disruption created by the 2008 crisis. Following the crisis, revenue resumed its upwards trajectory, as seen with broader economic recovery. The analysis will also consider variations of revenue generation per industry, as these also exhibit geographic characteristics.

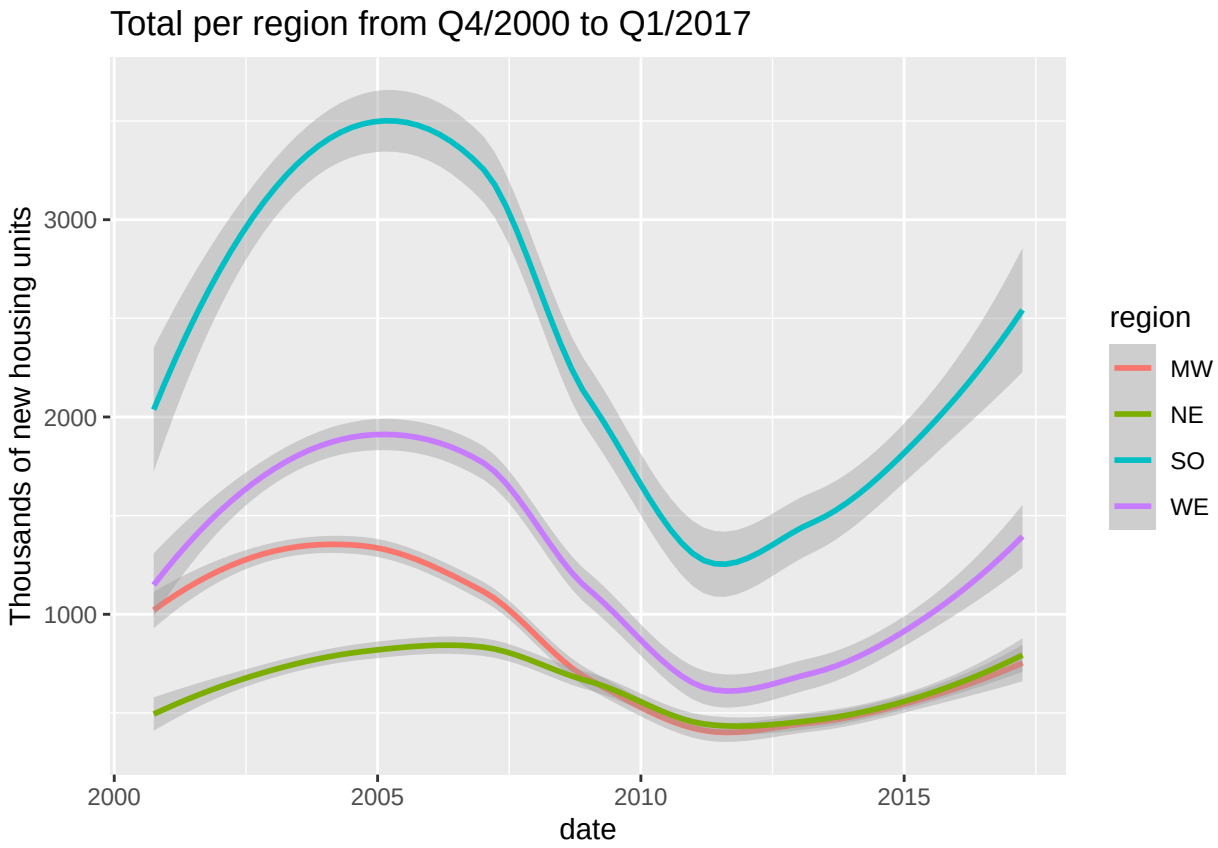
The regional graphs, however, show that new housing construction and offering vary along each individual regional average that appears not related to the broader revenue generation.

In order to understand what drives an increase and decrease of new housing, this analysis will classify trend as positive when the numbers are above average, and as negative when they are below in each distinct region.






```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Preparing data for analysis

Several steps are performed here:

(1) Aligning the time periods between income generation and house construction, allowing for a 1 year lag between the financial indicators being perceived and the construction to yield new houses.

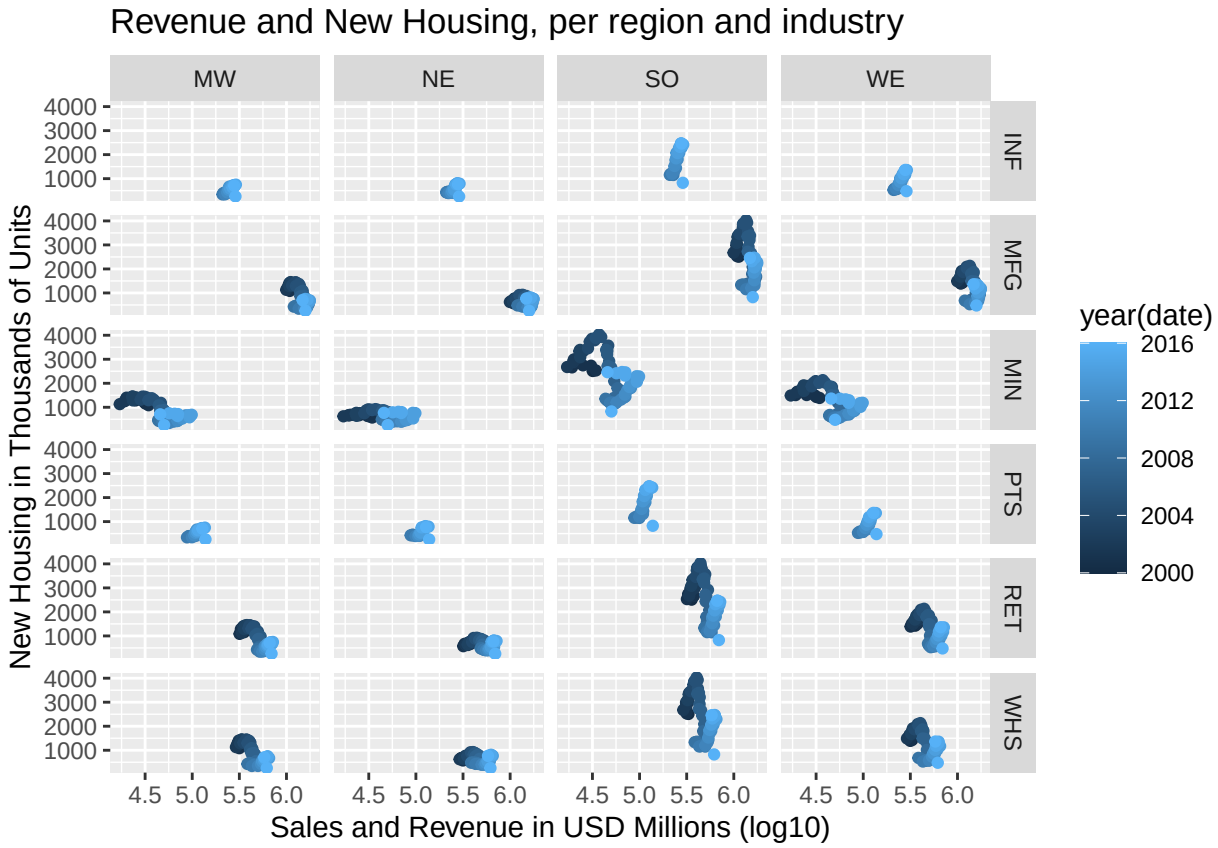
(2) Removing aggregated US National data points in order not to double count properties.

(3) Preparing regional averages, given that the variations are significant between regions.

(4) Introducing industry classification, given their regional variation.

Industry codes are:

Code	Description
MFG	All Manufacturing
MIN	All Mining
RET	All Retail Trade
WHS	All Wholesale Trade
INF	All Information
PTS	All Professional and Technical Services, Except Legal Services



Making predictions

This report aims to demonstrate that the regional trends of new house construction is lagging the sentiment of a positive economic position, as observed by financial reports on sales, invoicing and revenue at the national level.

In order to do that, the concept of positive or negative trend was introduced. Positive trend means that regional construction is above historical average, and negative trend means below average.

The models that appeared most appropriate for this exercise were GLM, KNN and Random Forest.

The ensemble of those models did not improve on the prediction made via Random Forest.

	Model	Accuracy
Accuracy	GLM	0.812
Accuracy1	RF	0.959
Accuracy2	KNN	0.840
Accuracy3	Ensemble	0.910

Conclusion

There is a reasonably high predictive power of house construction by US Region based on the past year financial performance of each industry. The maximum accuracy obtained with the Random Forest model was 0.959.

Armed with this knowledge, there is potential for better decision making and reduction of unintended, excessive housing supply.