Team members: Anuraag Motiwale, Harshal Gurjar, Aditya, Abhishek Singh.

Problem statement: To predict the GDP for a state for a given year based on number of job openings, literacy rate, employment rate and per capita income for a given state during given year.

Datasets used:

1) Jobs.csv
2) Education.csv
3) Econ_state

We aggregated the data state wise and year wise for all states, for all years across 2011 to 2016.

This way we ended up with 600 permutation for state and year. We computed the corresponding values for these permutations to come up with a dataset ready to be fitted in a linear model.

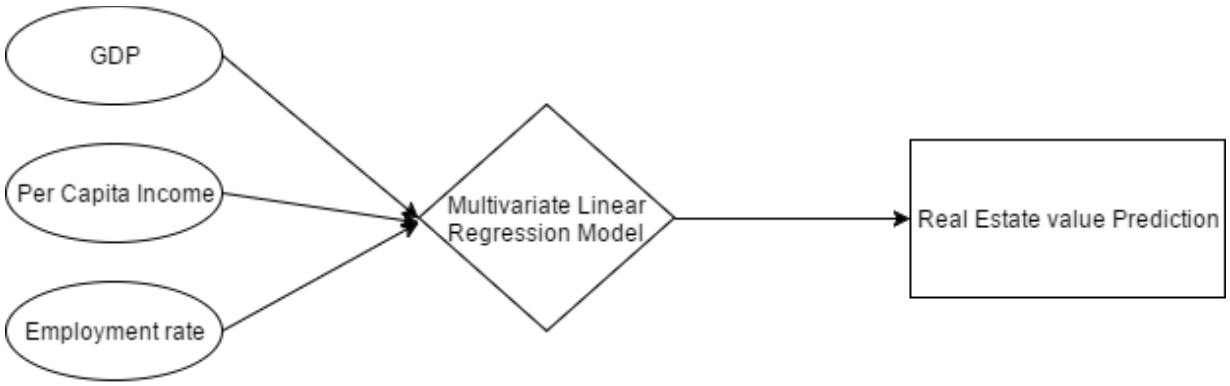
We trained the model using the 2011,2012 and 2013 data and made predictions on 2014 and 2015 values.

The results exhibited promising results with an R squared value of 0.9921. This model performed really well in predicting the GDP of a state given the values of job openings, literacy rate, employment rate and per capita income of the given state. GDP showed a strong linear correlation with literacy rates and number of job openings, strong anti-correlation with unemployment rate.

This model can used to predict the GDP of a state. Summary of the model is as following:

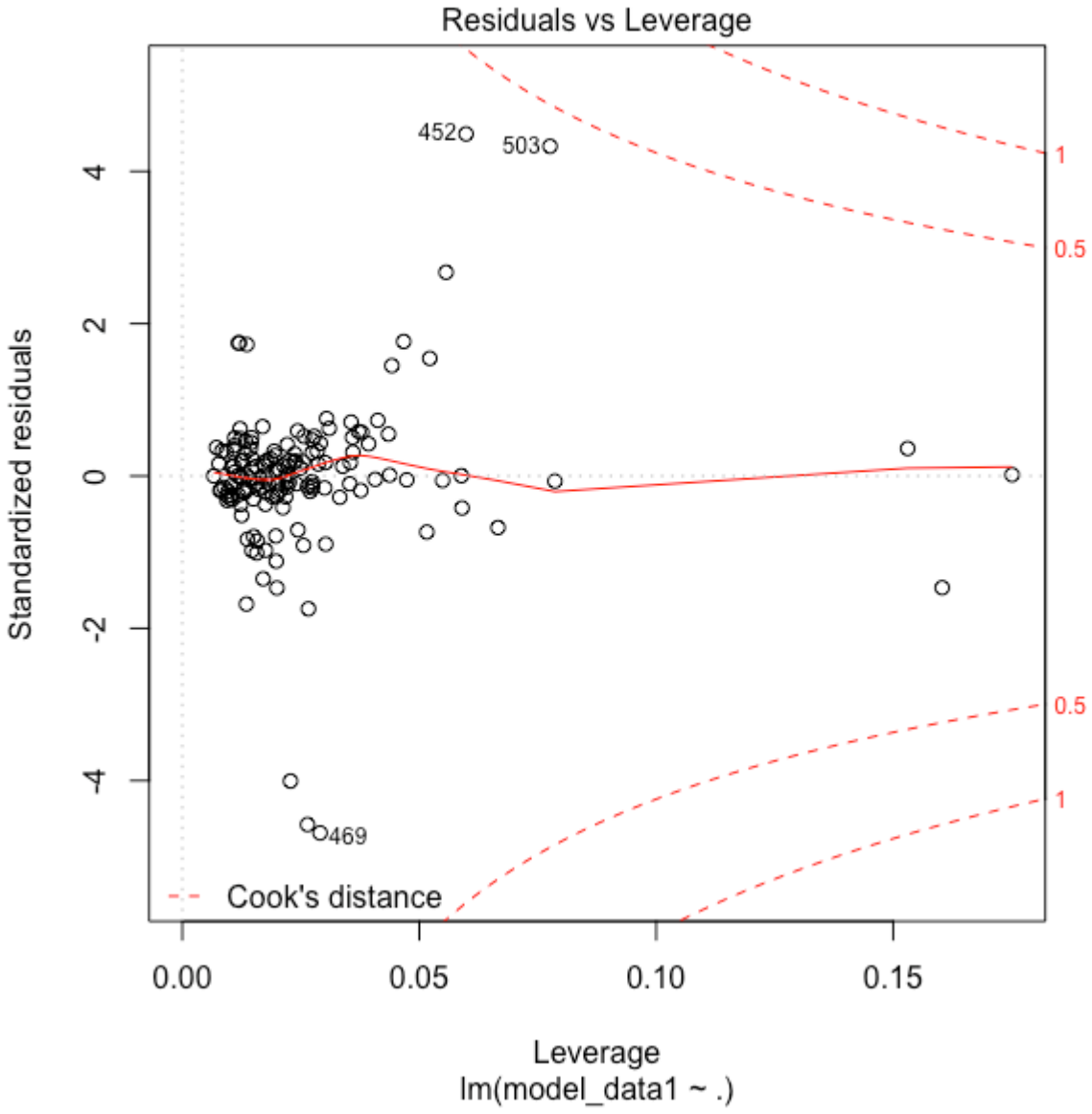
```
Call:
lm(formula = model_data1 ~ ., data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-176199   -8150      27   12271  166129

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.928e+04  4.856e+04  -1.015    0.312
model_data2  1.191e-03  9.649e-06 123.384   <2e-16 ***
model_data3 -6.630e+02  1.943e+03  -0.341    0.733
model_data4  5.252e+04  5.113e+04   1.027    0.306
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 38160 on 149 degrees of freedom
Multiple R-squared:  0.9913,    Adjusted R-squared:  0.9911
F-statistic:  5637 on 3 and 149 DF,  p-value: < 2.2e-16
```

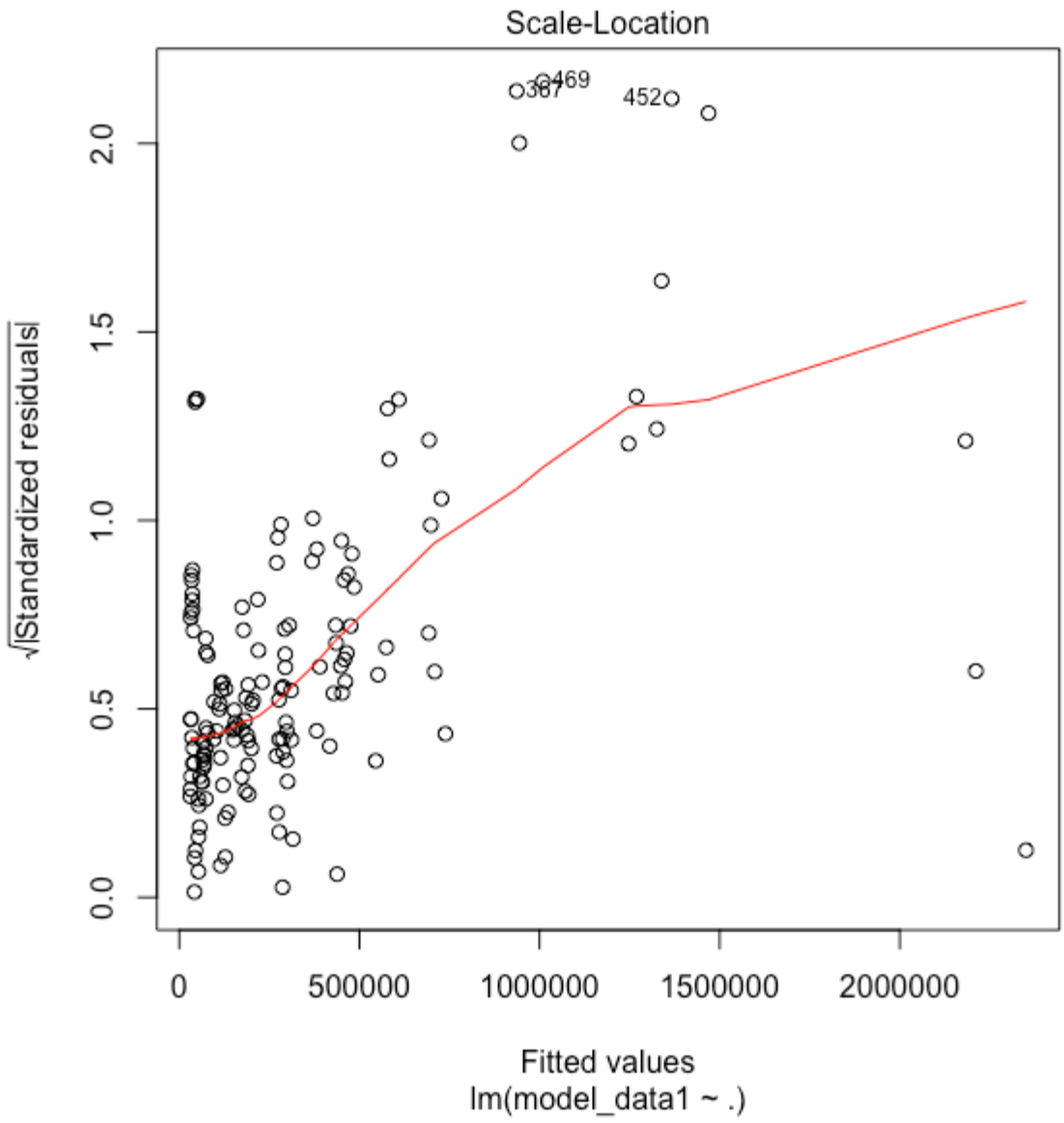
Model overview:

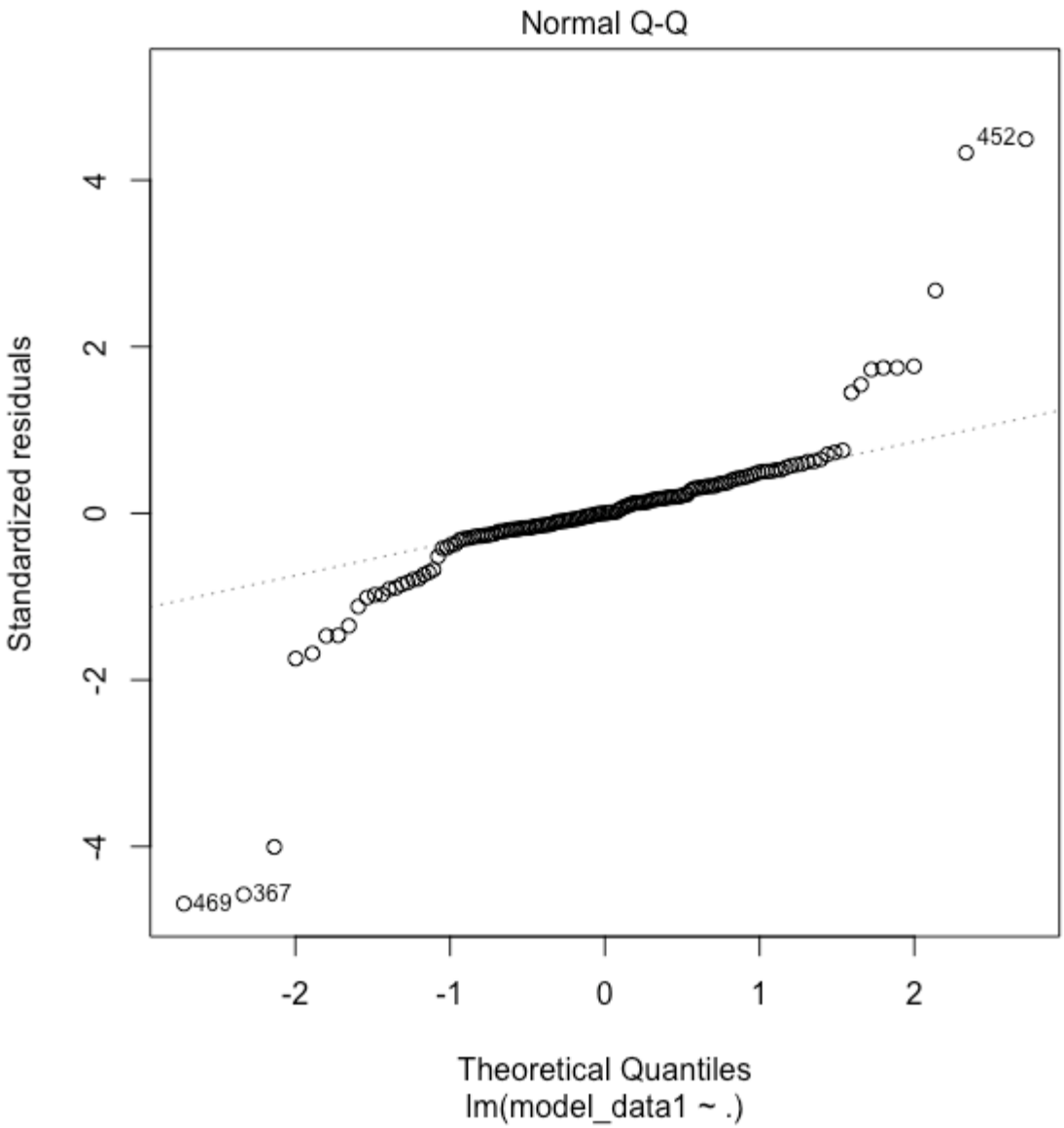


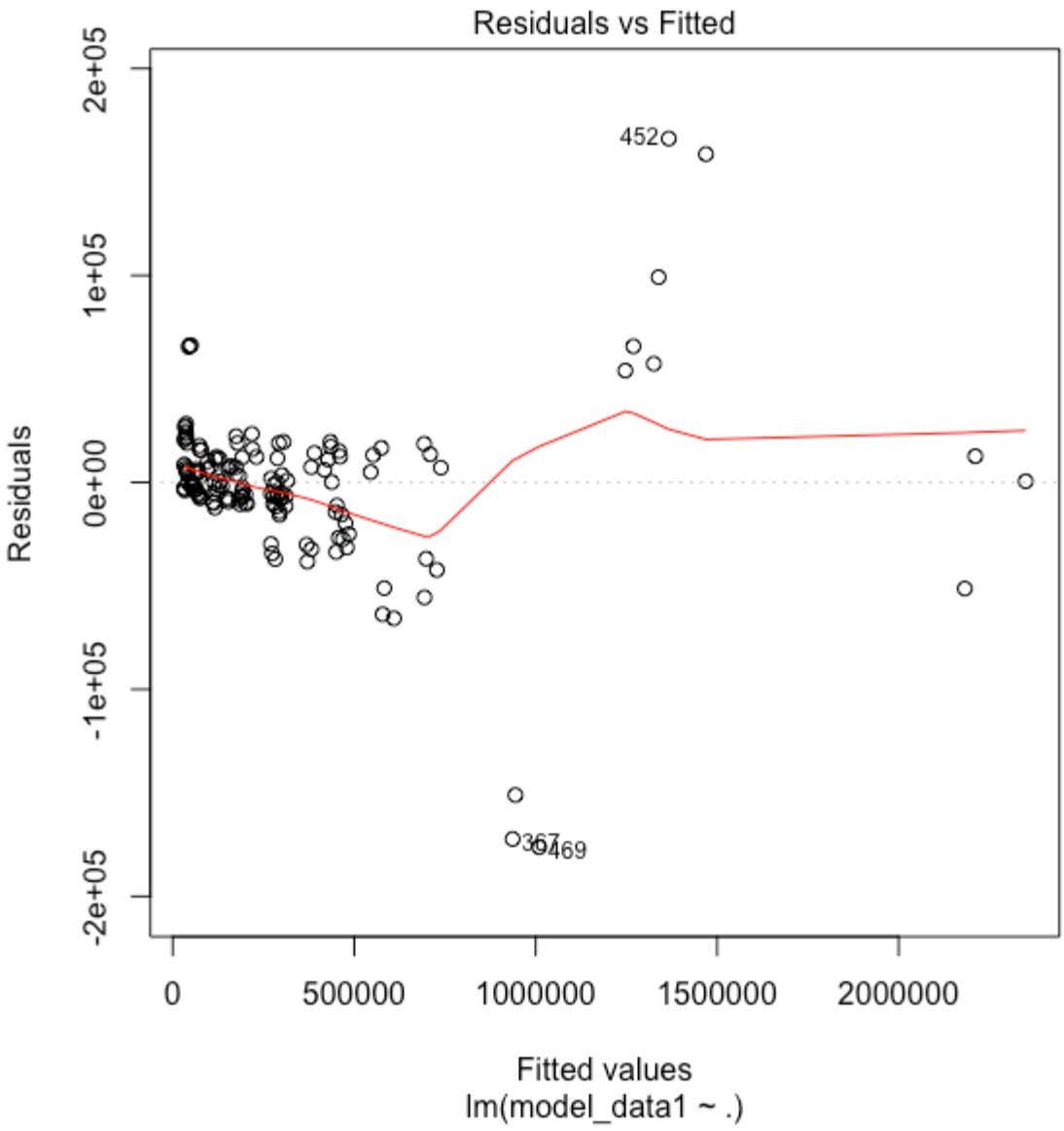
Linear Regression model for predicting Real Estate values

Model performance summary graphs:

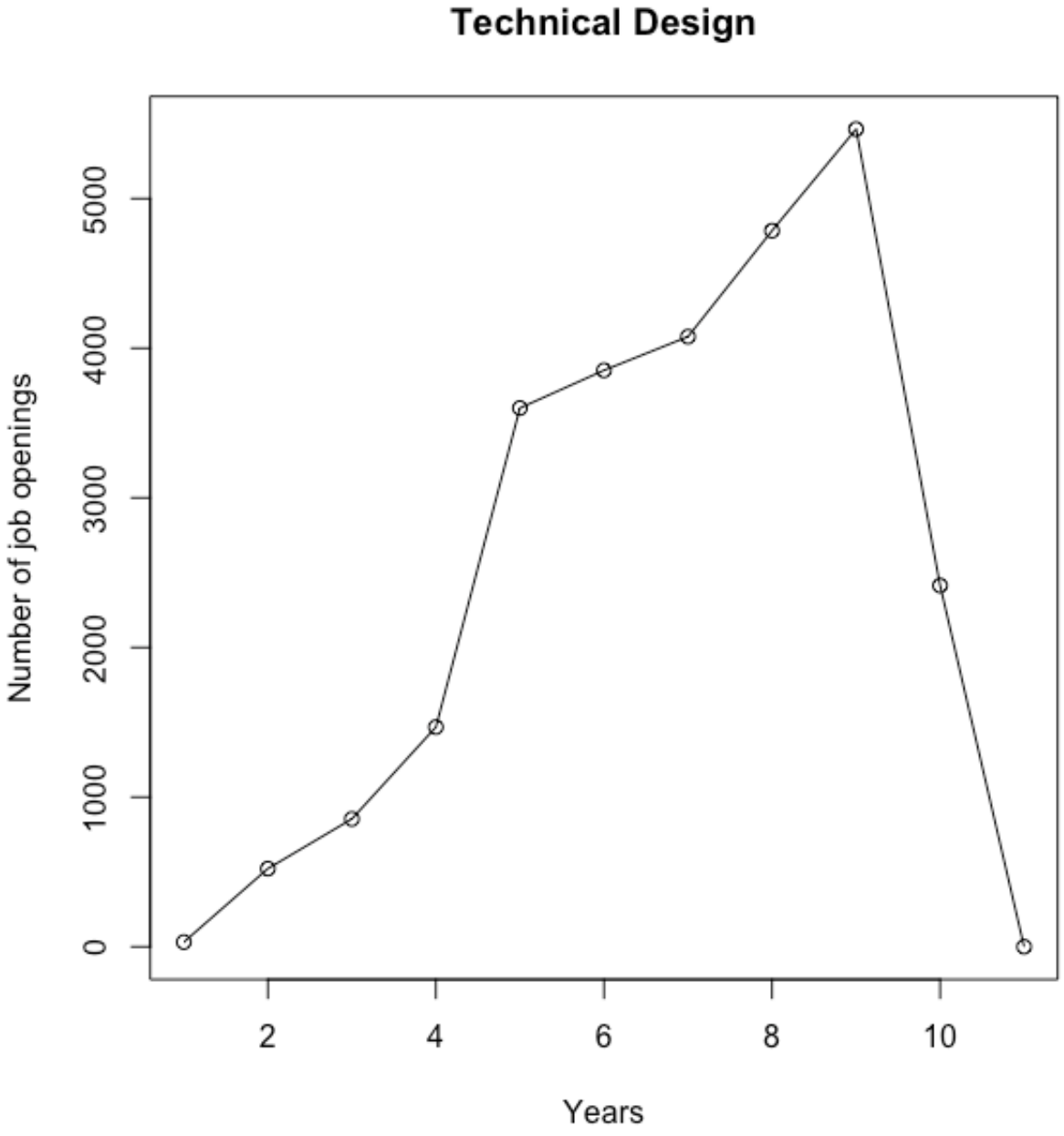


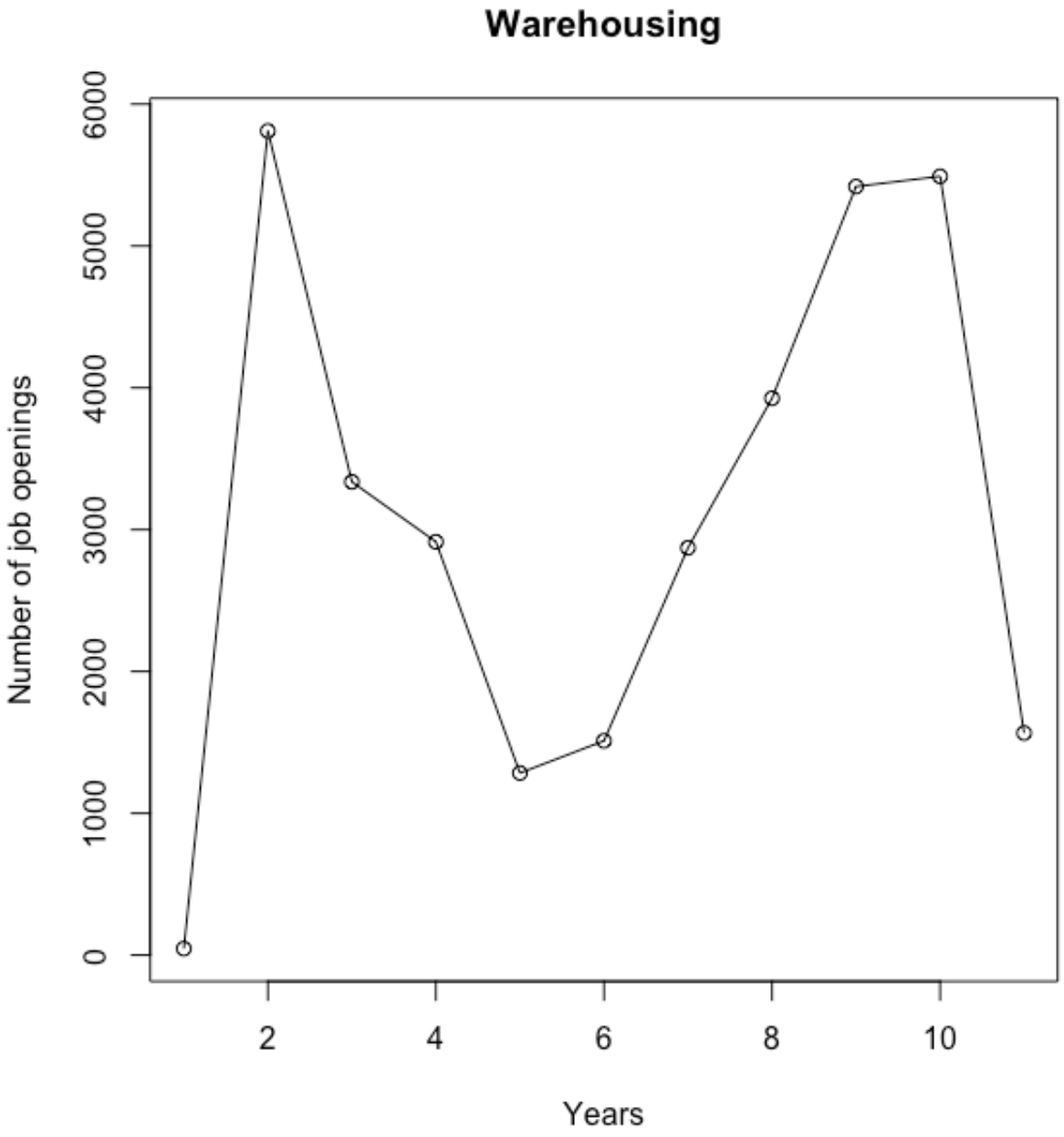

Scale-Location
469
387
452
2.0
1.5
1.0
0.5
0.0
√|Standardized residuals|
0
500000
1000000
1500000
2000000
Fitted values
lm(model_data1 ~ .)

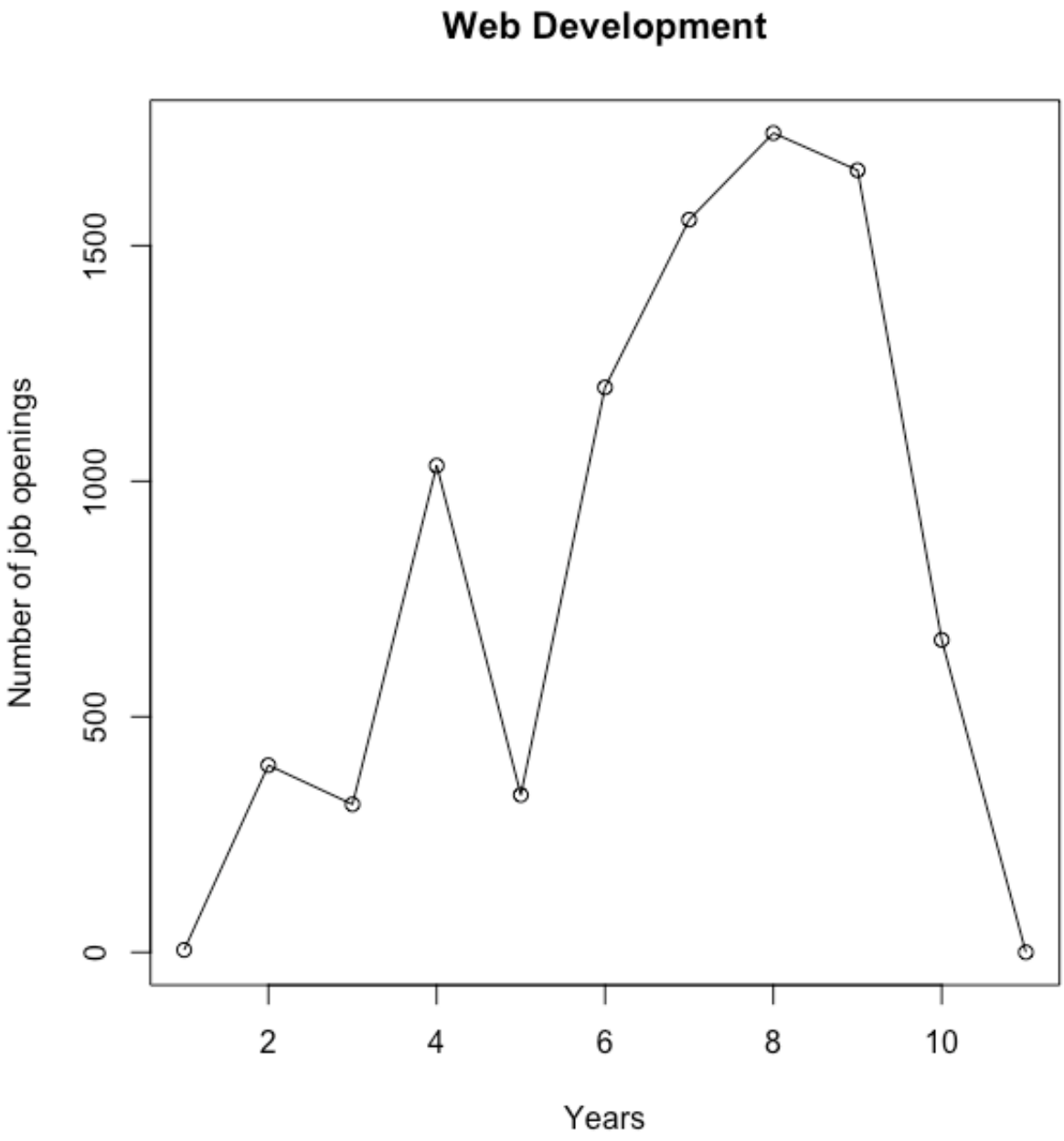

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(model_data1 ~ .)
452
469
367
4
2
0
-2
-4
-2
-1
0
1
2

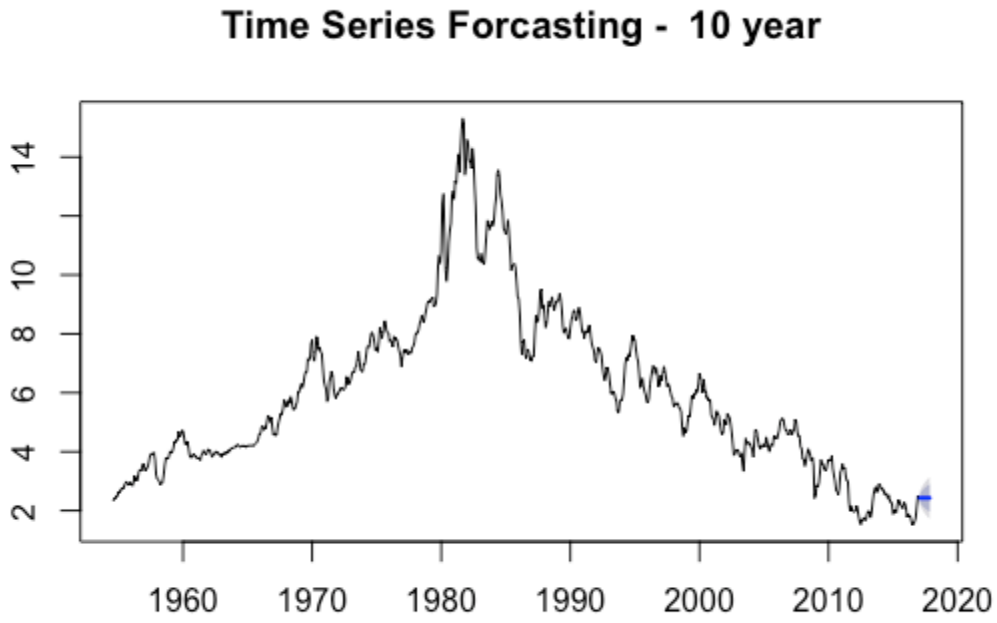

Residuals vs Fitted
Residuals
Fitted values
lm(model_data1 ~ .)
2e+05
1e+05
0e+00
-1e+05
-2e+05
0
500000
1000000
1500000
2000000
452
367
469

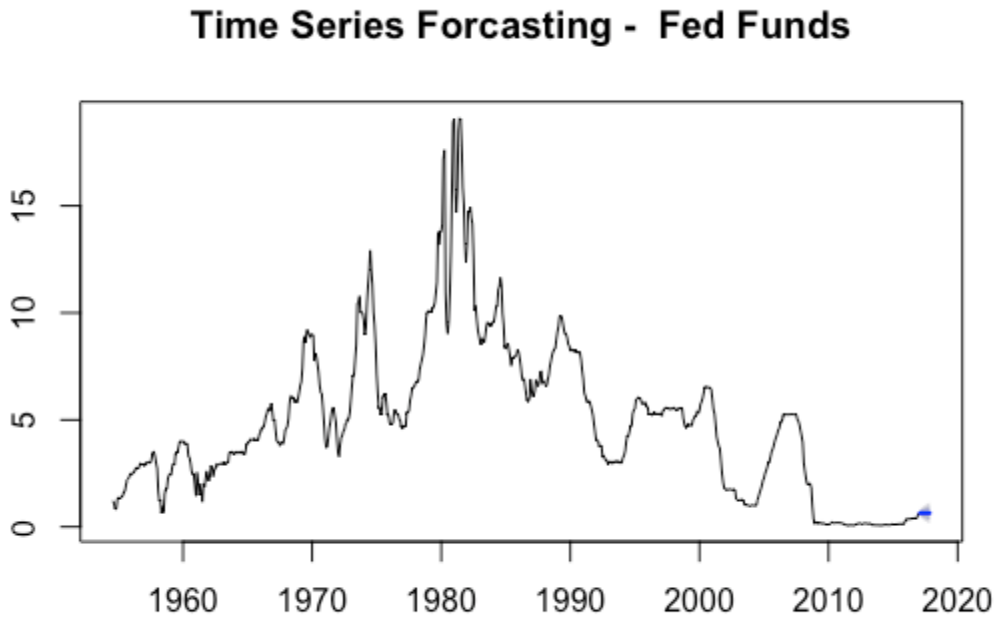
We analyzed the trends present in the categories of the job openings over the years. We found that the number of job openings were affected significantly after the great recession of 2008. While, there was steady growth observed in the number of job openings in some categories, but the job openings in the category of “Manufacturing and Operations” were affected.

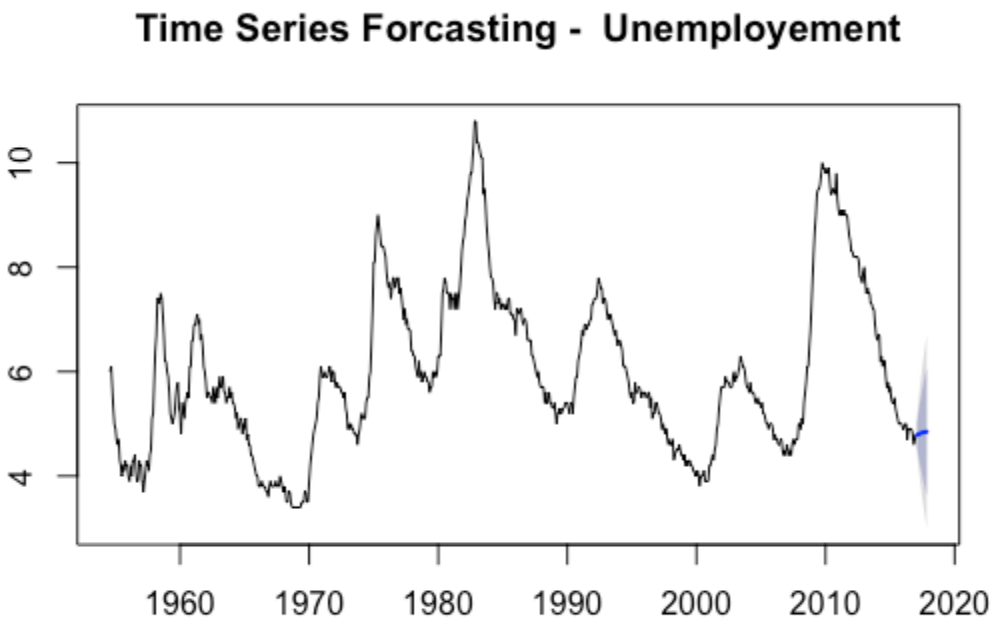


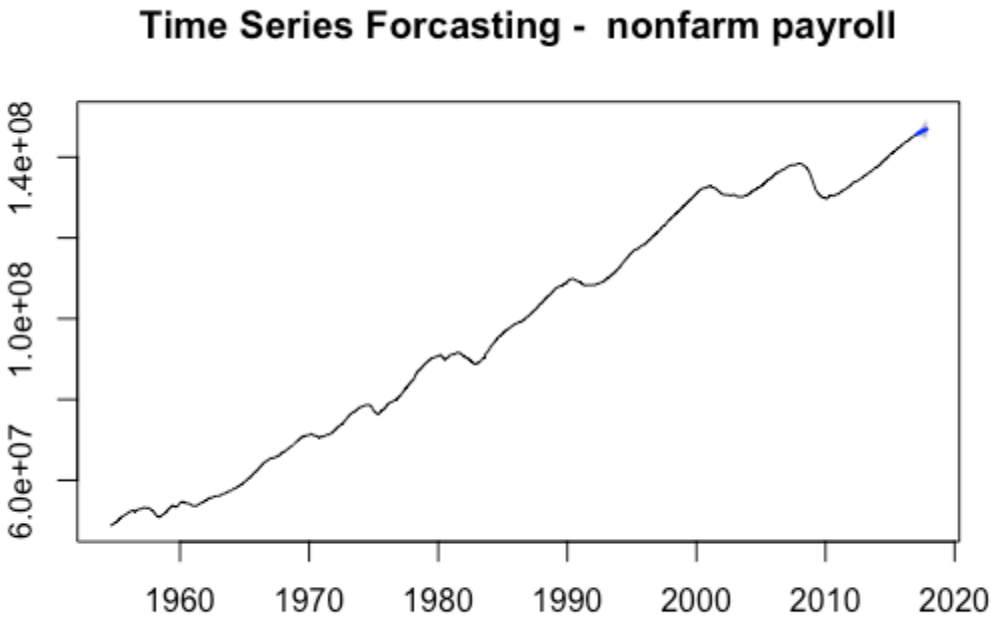

Warehousing
Number of job openings
0
1000
2000
3000
4000
5000
6000
2
4
6
8
10
Years

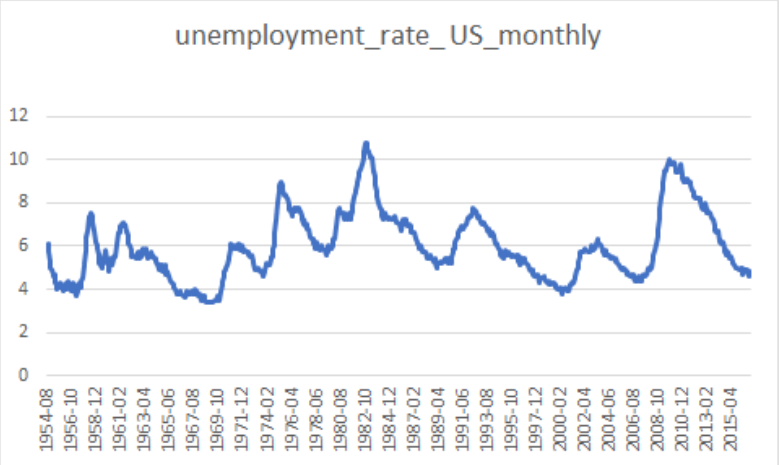

Web Development
Number of job openings
0
500
1000
1500
2
4
6
8
10
Years

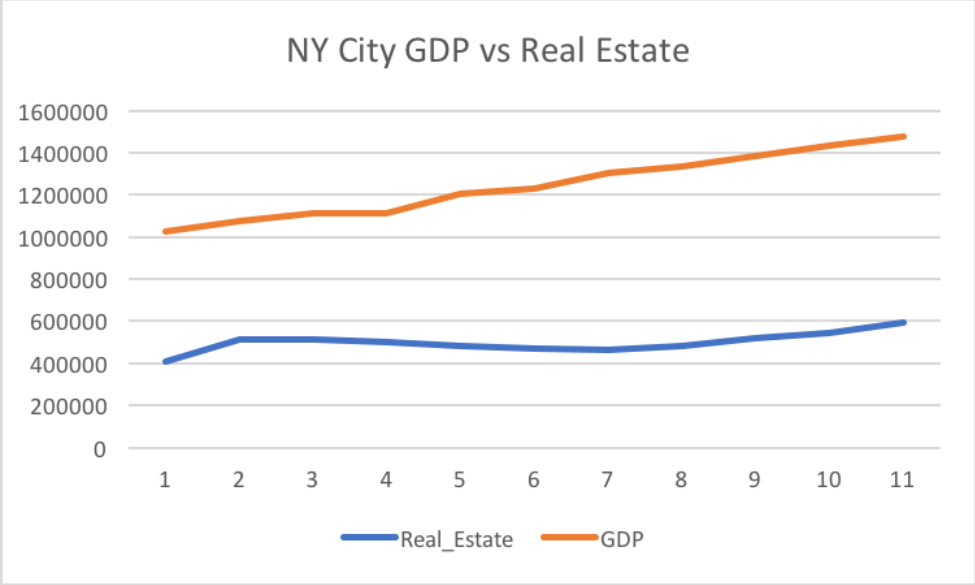

Time Series Forcasting - 10 year
2
4
6
8
10
14
1960
1970
1980
1990
2000
2010
2020


Time Series Forcasting - Fed Funds
0
5
10
15
1960
1970
1980
1990
2000
2010
2020


Time Series Forcasting - Unemployement
4
6
8
10
1960
1970
1980
1990
2000
2010
2020


Time Series Forcasting - nonfarm payroll
6.0e+07
1.0e+08
1.4e+08
1960
1970
1980
1990
2000
2010
2020


unemployment_rate_US_monthly
12
10
8
6
4
2
0
1954-08
1956-10
1958-12
1961-02
1963-04
1965-06
1967-08
1969-10
1971-12
1974-02
1976-04
1978-06
1980-08
1982-10
1984-12
1987-02
1989-04
1991-06
1993-08
1995-10
1997-12
2000-02
2002-04
2004-06
2006-08
2008-10
2010-12
2013-02
2015-04


NY City GDP vs Real Estate
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
1
2
3
4
5
6
7
8
9
10
11
Real_Estate
GDP