REGRESSION IN FINANCIAL ANALYTICS

MITHUN MEENAKSHI S, UMIP7194

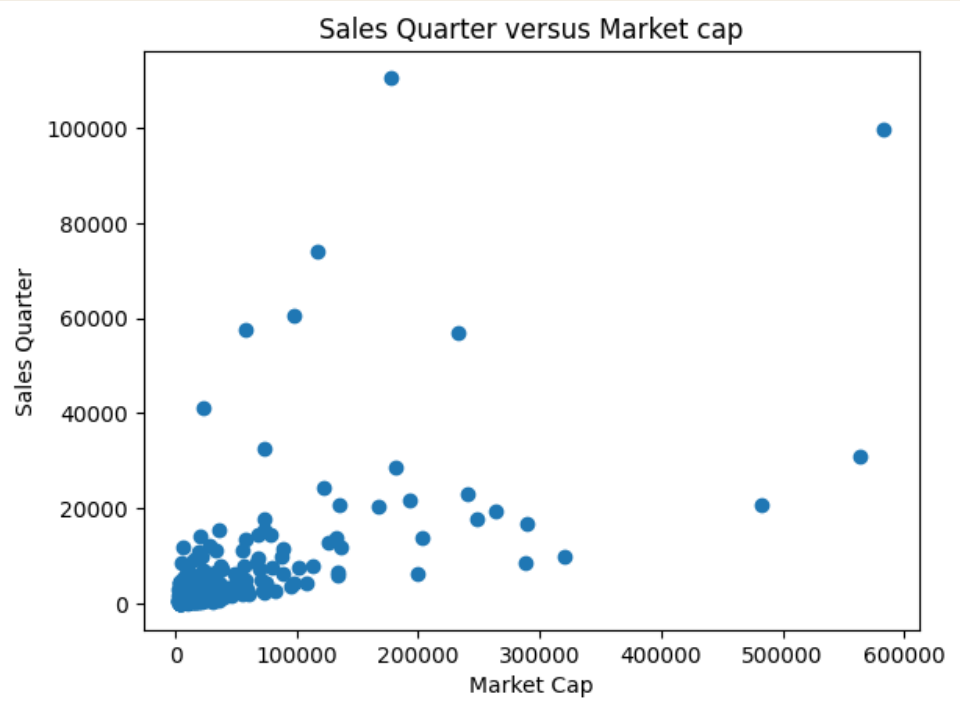
TABLE OF CONTENTS

DATA CLEANING

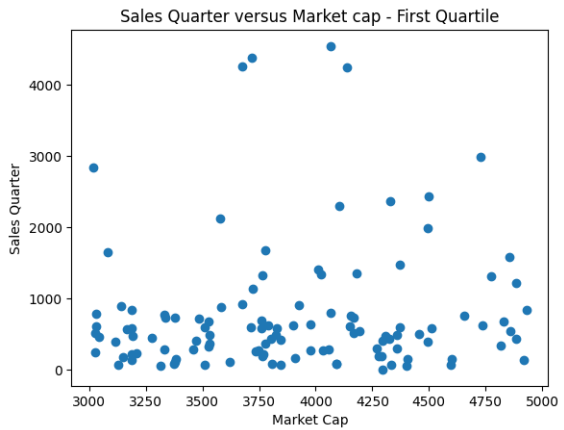
- Preliminary data cleaning was conducted using Microsoft Excel
- Fixed column names in snake format
- Used sorting and filtering to look up empty and NaN values
- Removed all rows with Nan and empty values
- Did not treat for outliers due to nature of data.

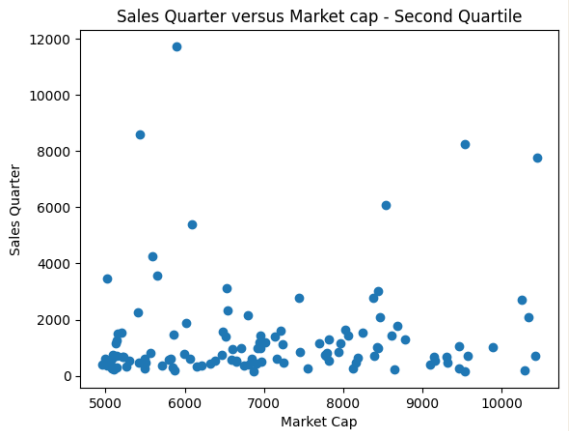
DATA EXPLORATION

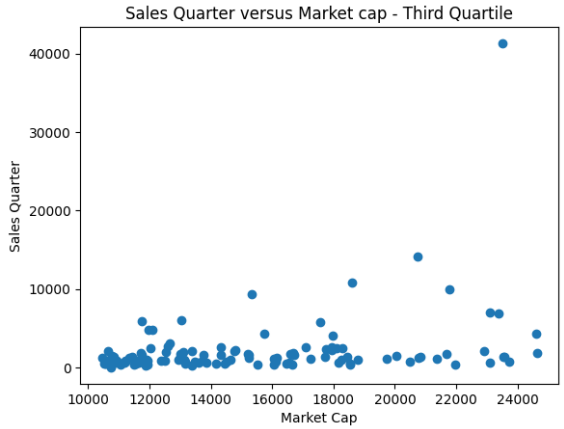
- Basic packages and modules were imported.
- Used basic functions like describe(), info(), shape to study the structure of the dataset
- Shape: (459, 3)
- Size: 1377

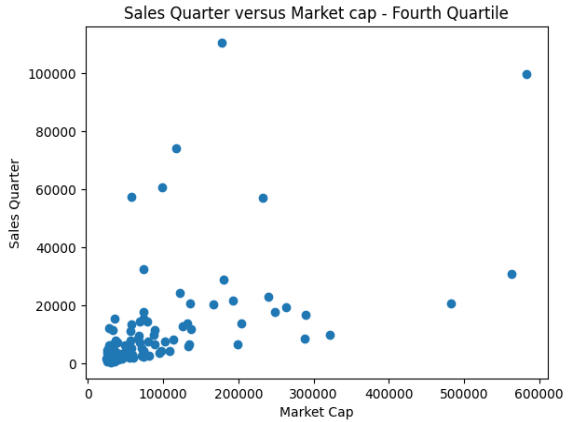


DATA EXPLORATION



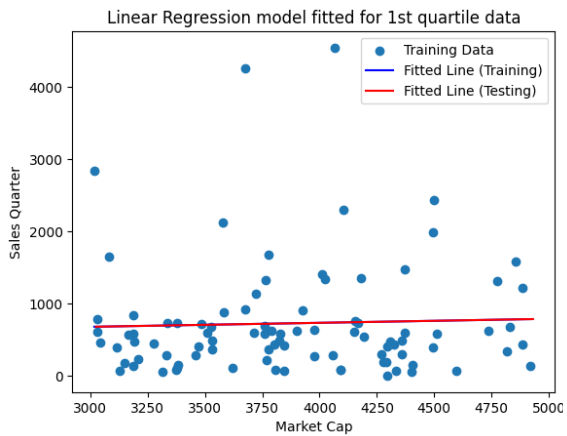


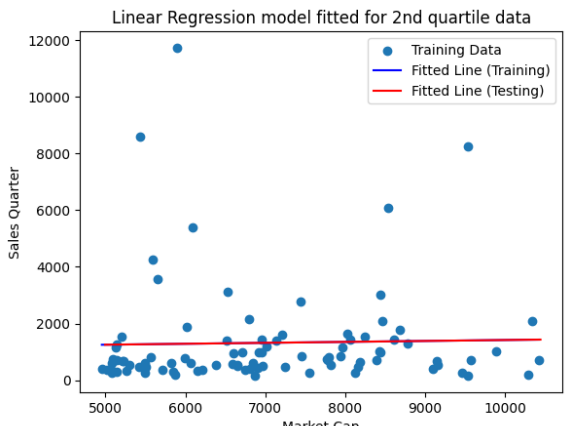


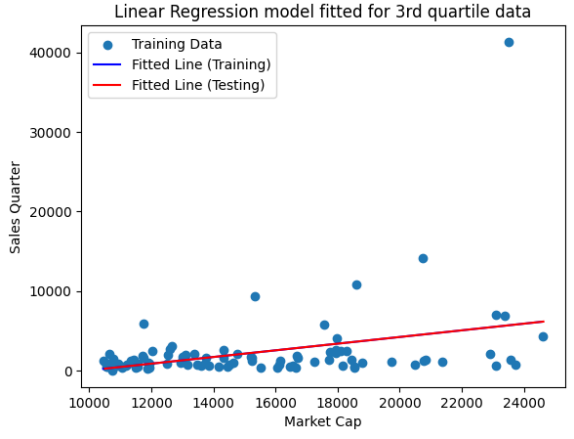


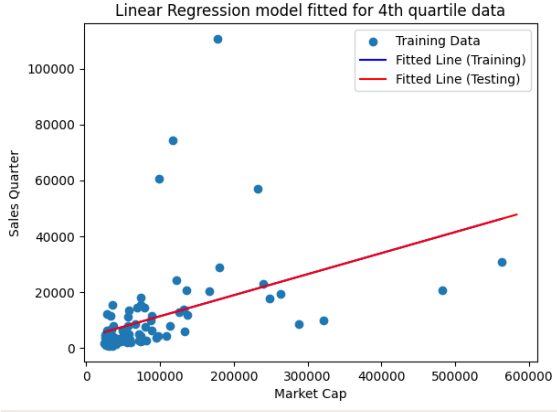
- To study the dataset, the dataframe was split up into 4 quartiles, and similar plots were plotted.
- A linear relationship can be observed in plot 1, 2 and 3.

REGRESSION



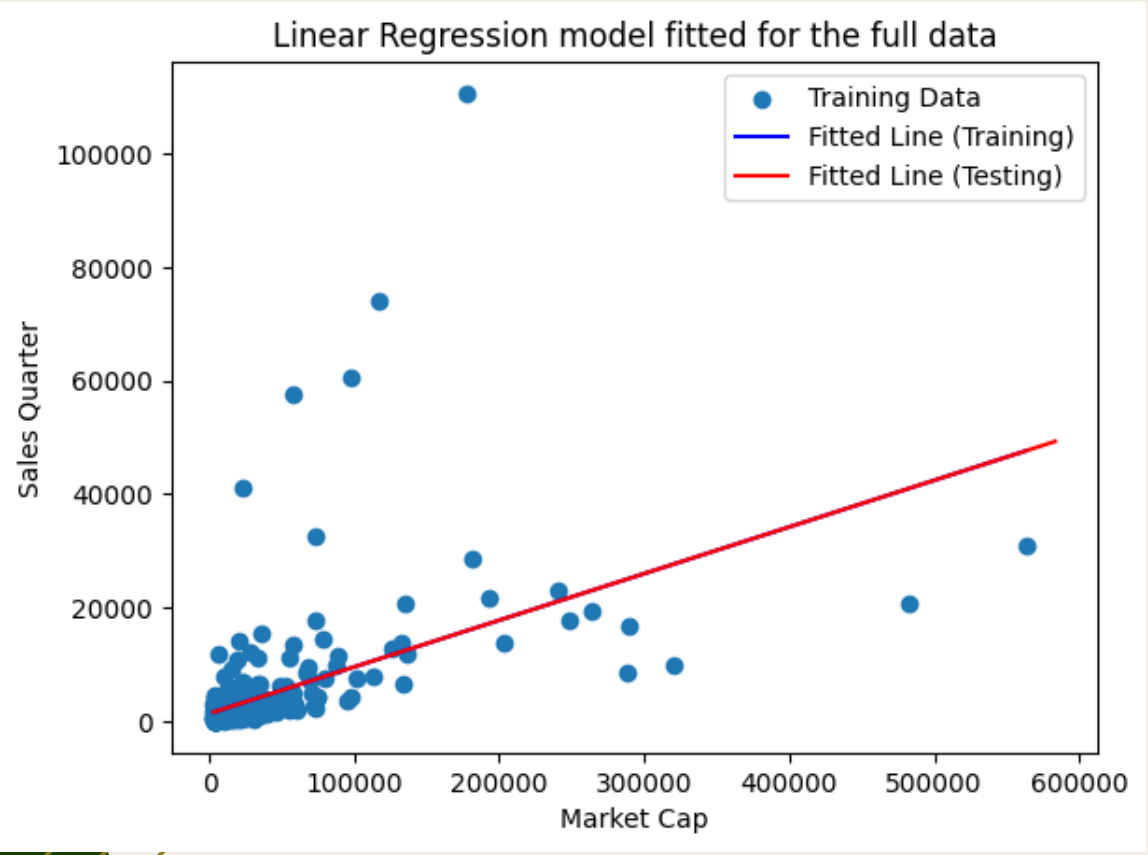






- Linear regression models were fit for all 4 quartiles separately, and yielded a good fit
- However, the 3rd quartile has a negative intercept which could have been avoided using outlier elimination
- Outlier elimination was not conducted because of nature of data.

REGRESSION



- This plot shows the linear regression model fit for the whole data.
- As we can see, we cannot draw much inferences due to the high density of points on the lower left corner.

CONCLUSION

- Preliminary data cleaning was conducted using Microsoft Excel
- Structure of data was explored using basic functions
- Linear regression model was fit on whole data as well as the four quartiles
- Outlier elimination could give better results, but its not advisable to do so in financial data.


slidesgo