Machine Learning

Weekly Project Report

Quantcats

Prachee Javiya AU1841032
Kaushal Patil AU1841040
Arpitsinh Vaghela AU1841034
Vrunda Gadesha AU2049007

Tasks Performed: Week 2

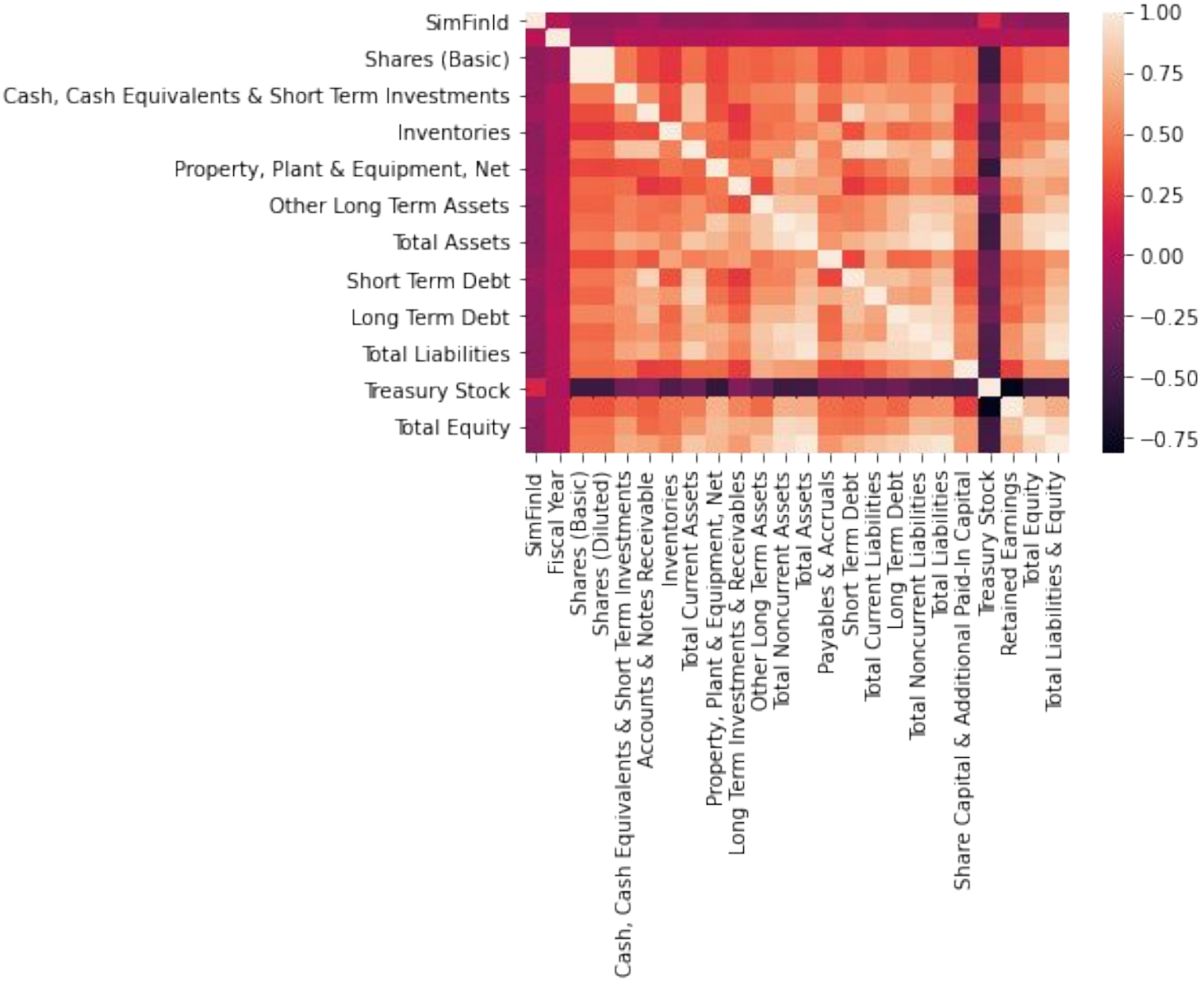
01 After finalizing the topic, Stock Classifier, we finalised the dataset - link.

02 Initial data analysis with the dataset

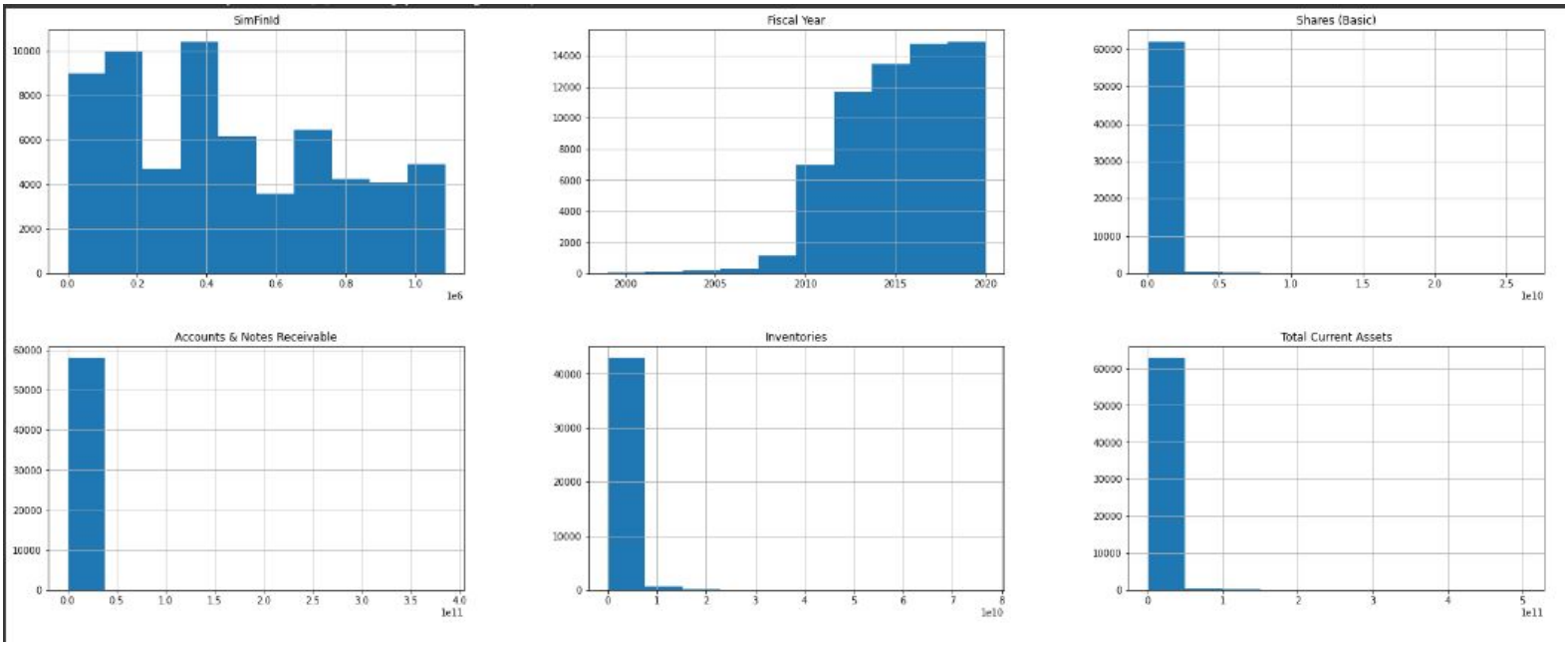
03 Tried to perform basic exploratory dataset analysis by doing the following:

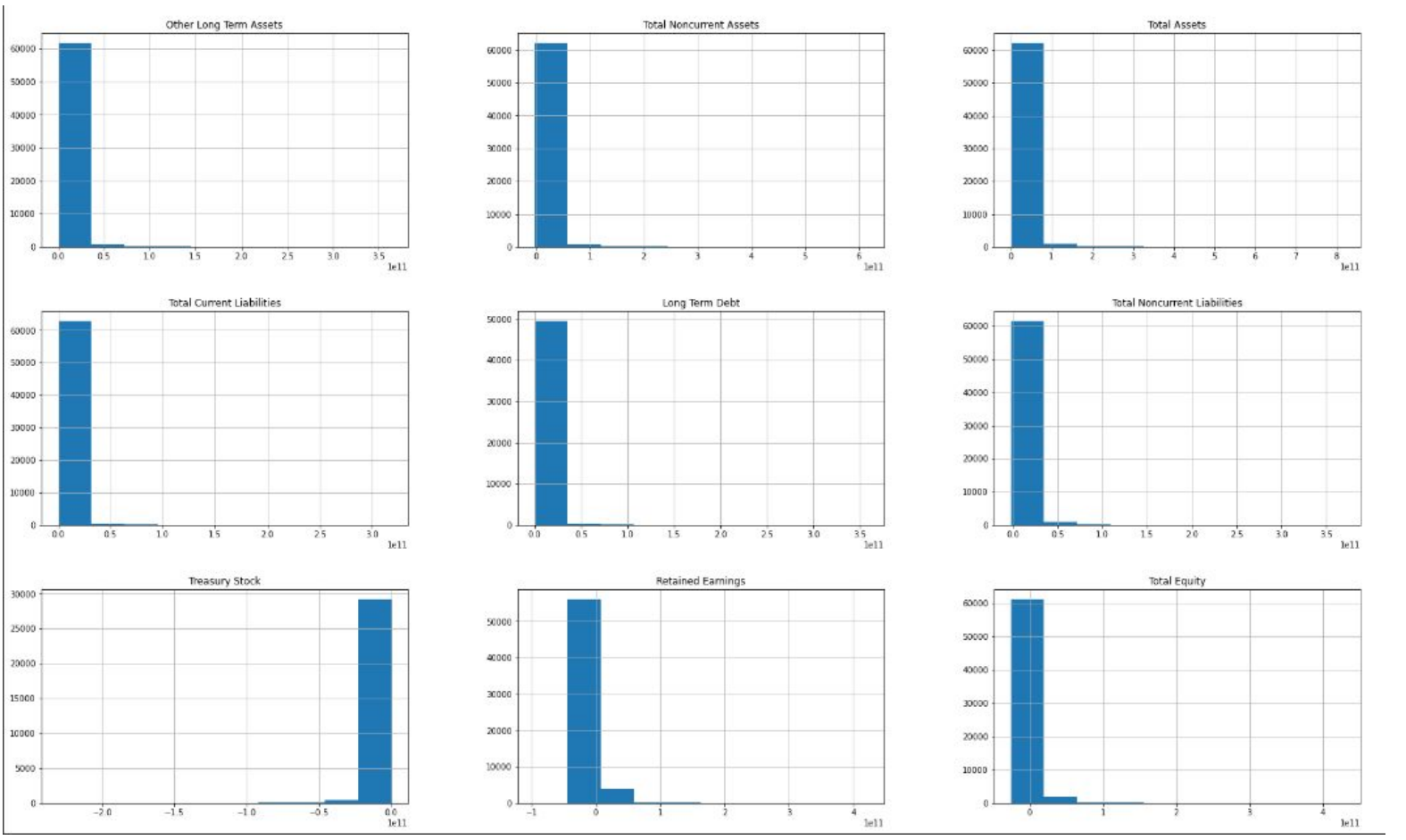
- Load data, create data table, number of fields(30), correlation
- Total number of data points: 63,547
- Plotted heatmap of correlation using seaborn
- Plotted Histogram for all columns to see how data is distributed for different fields

Tasks Performed: Week 2

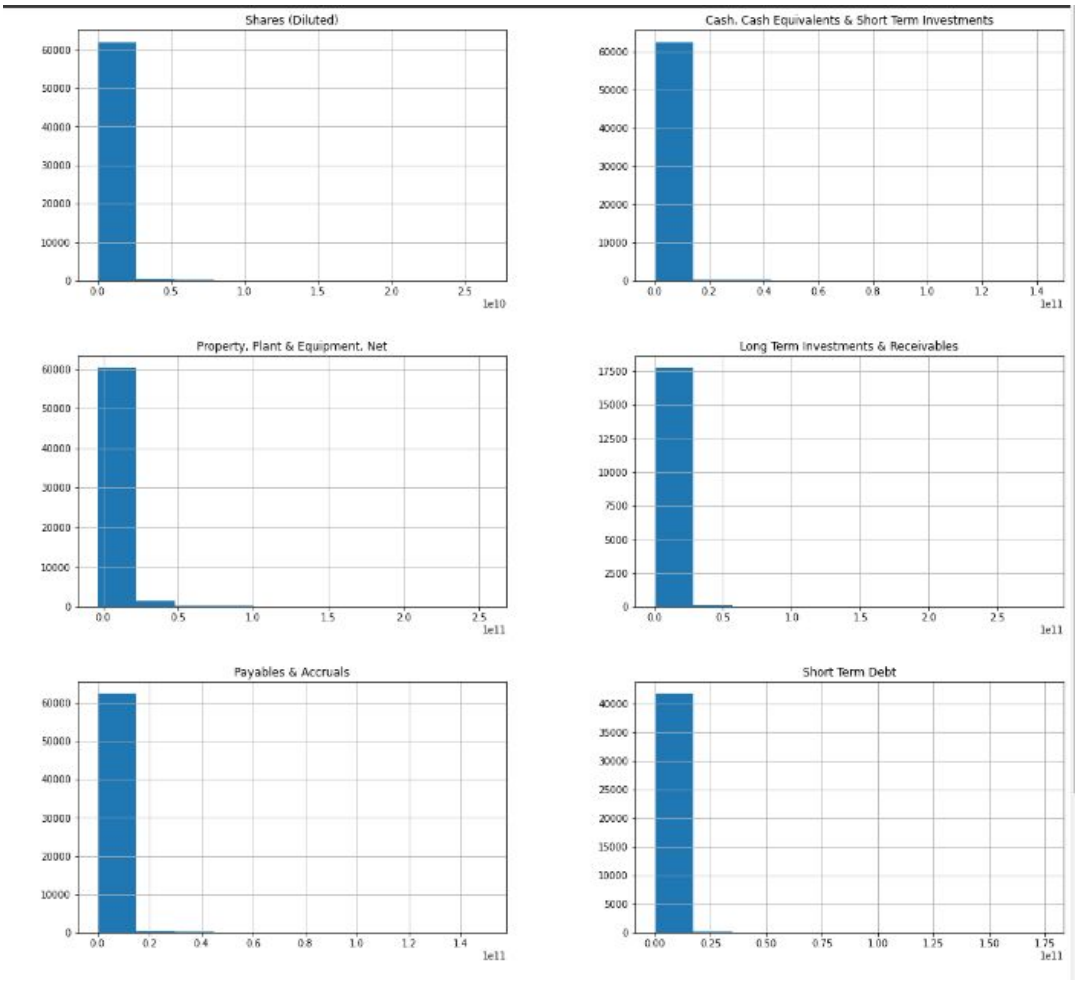


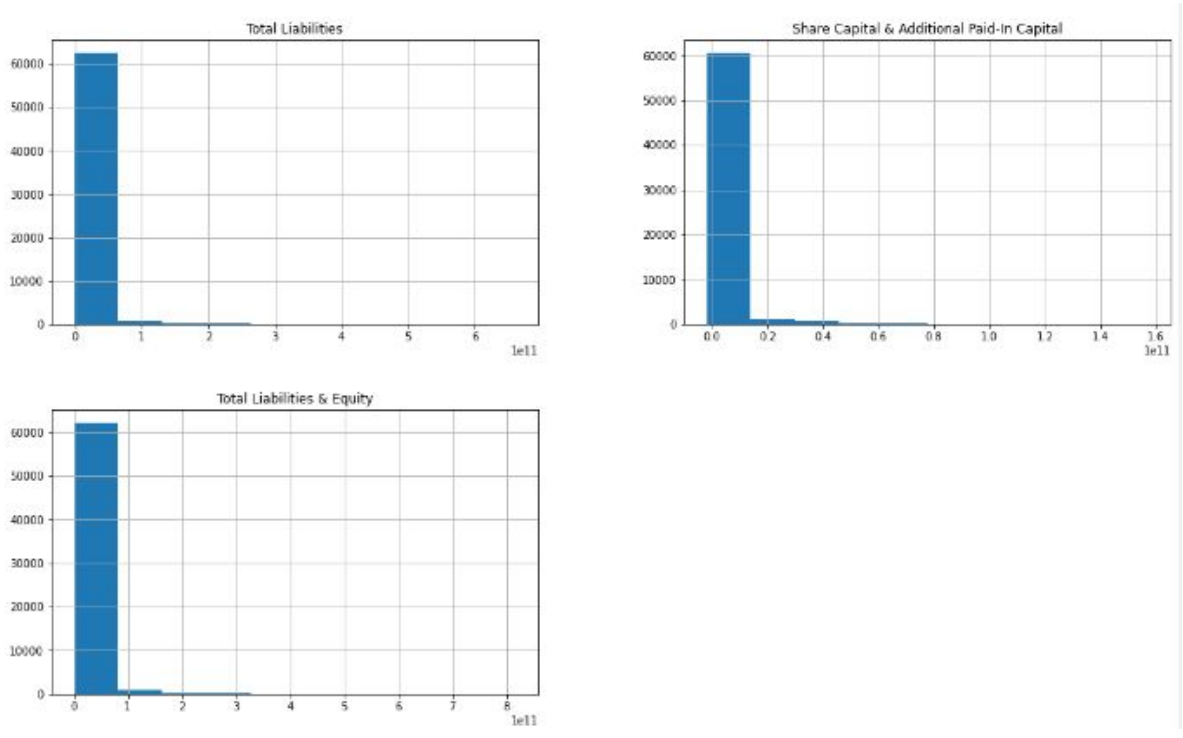
Tasks Performed: Week 2





Tasks Performed: Week 2





Outcomes

Issues discussed:

Dataset Manipulation, Increasing Features with High Impact

- We can probably use other datasets from https://simfin.com/data/bulk
- We can even combine technical indicators from daily price datasets.

Labelling the Dataset

- Labelling the data based on stock prices on fundamental results declaration date and price on the date from 3 months of results declaration date.

Requires us to write a script that takes the date from the Declare Date row of our primary data set (https://simfin.com/data/bulk)(select dataset: Balance Sheet Quarterly) and fetches price for that date and three months ahead from US daily share prices from (https://simfin.com/data/bulk)(select dataset: Share Prices) for that particular Ticker.

If future price greatest than X% label BUY, if price less than X% SELL, if price between X & -X% HOLD/ Do Nothing.

Upcoming Week

01 Writing script for data labelling

02 Performing other advanced exploratory data set analysis methods

03 Testing the created dataset on fairly simple multilabel classification models such as Logistic Regression, LDA, & Random Forests(Decision Trees),