Financial Risk Analysis

Team 9

Balamanikandan Gopalakrishnan

Rachan Hegde

Agenda

- Problem Statement

- Data Set

- Approach

- Prediction – Impacted Stocks, VAR

Problem Statement

- Estimate the VAR (Value at Risk) associated with the given portfolio of stocks for a specific time period. Also provide insights on the stocks that are most impacted during this time frame

Dataset

- Portfolio of 3000 Stocks
  - Open, Close, High, Low, Volume and Adj Close
  - These values are captured on a daily basis
  - Market factors (S&P 500, NASDAQ, Bonds, Oil)
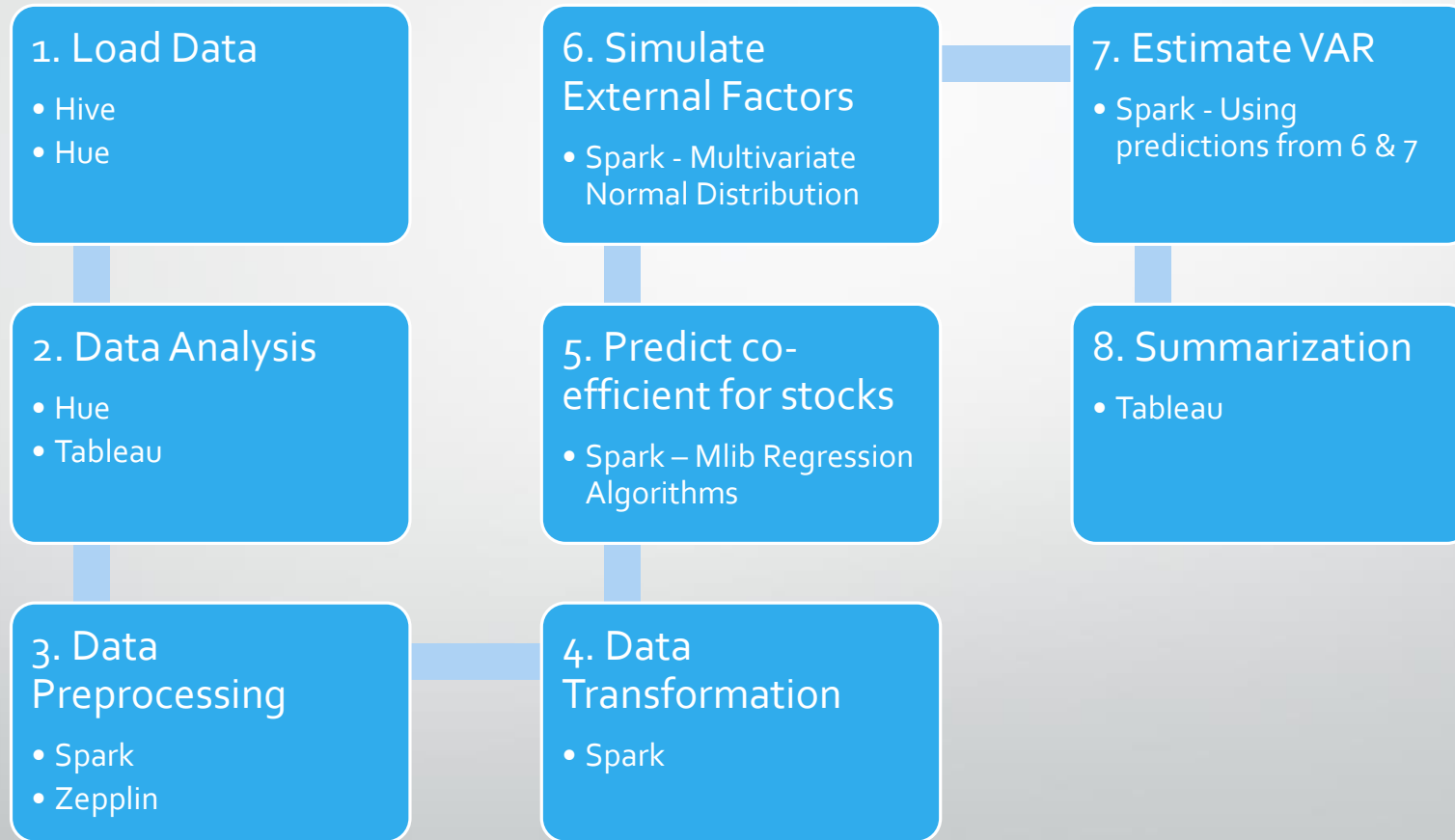 - ~ 9 Million Records

Additional Information

- The value of stocks are impacted by external factors (S&P 500, NASDAQ, Bonds, Oil)

- Financial Risk Analysis

  - Estimate the factor co-efficient associated with each of the stock by using the historical data

  - Using Monte Carlo simulation, estimate these external factors

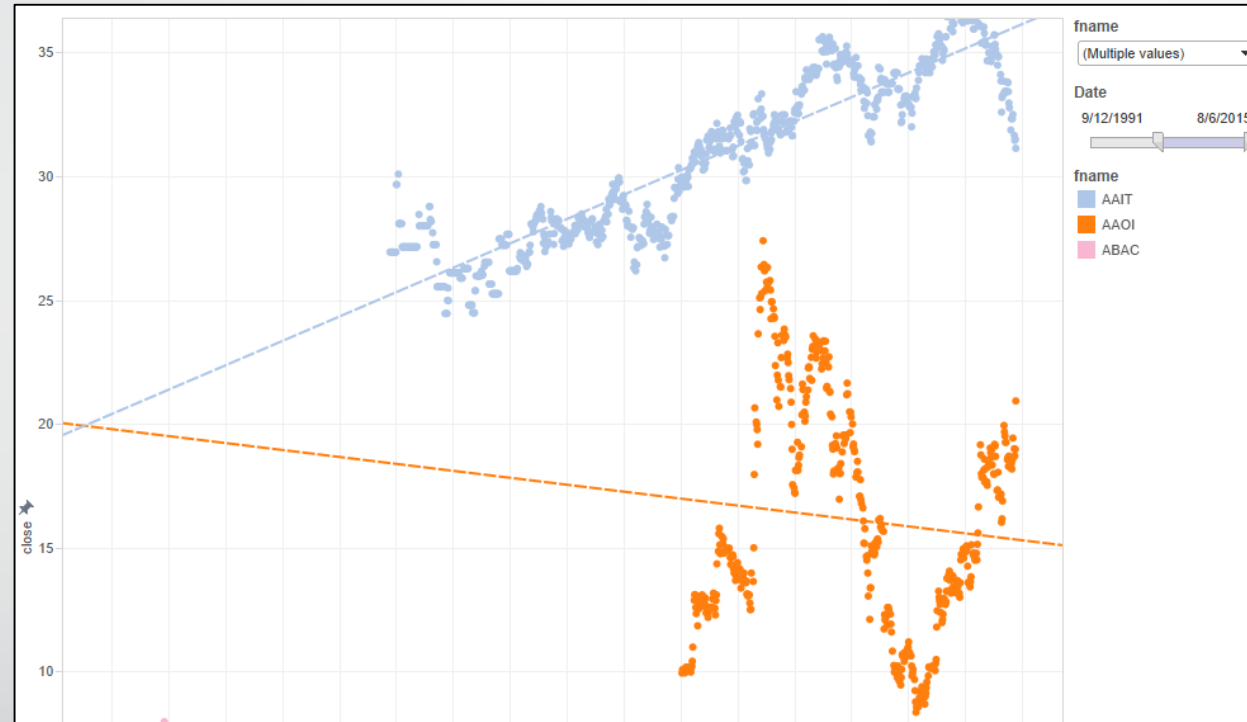 - Compute the VAR using estimated external factors and stock co-efficient

Environment Used

- Amazon EMR v3.8
    - 1 Master mx3 large instances
    - 4 Slaves mx3 large instances
- Spark 1.3.1
- Hue 0.17
- Hive
- Apache Zeppelin Bootstrapped
- Tableau

Approach

1. Load Data
- Hive
- Hue

2. Data Analysis
- Hue
- Tableau

3. Data Preprocessing
- Spark
- Zepplin

4. Data Transformation
- Spark

5. Predict co-efficient for stocks
- Spark – Mlib Regression Algorithms

6. Simulate External Factors
- Spark - Multivariate Normal Distribution

7. Estimate VAR
- Spark - Using predictions from 6 & 7

8. Summarization
- Tableau

Data Analysis – Missing Data Handling

- Note: The missing data was filled with observations from nearest neighbor as latest records are more relevant in time series data that using mean.

Data Analysis – Correlation
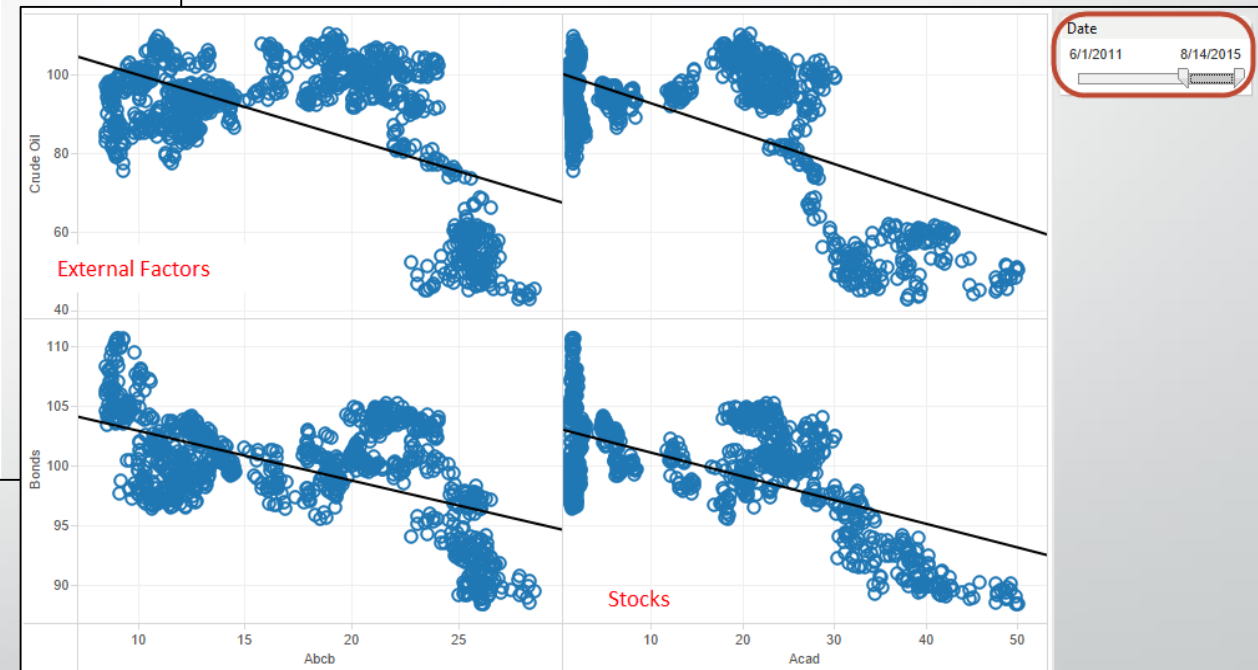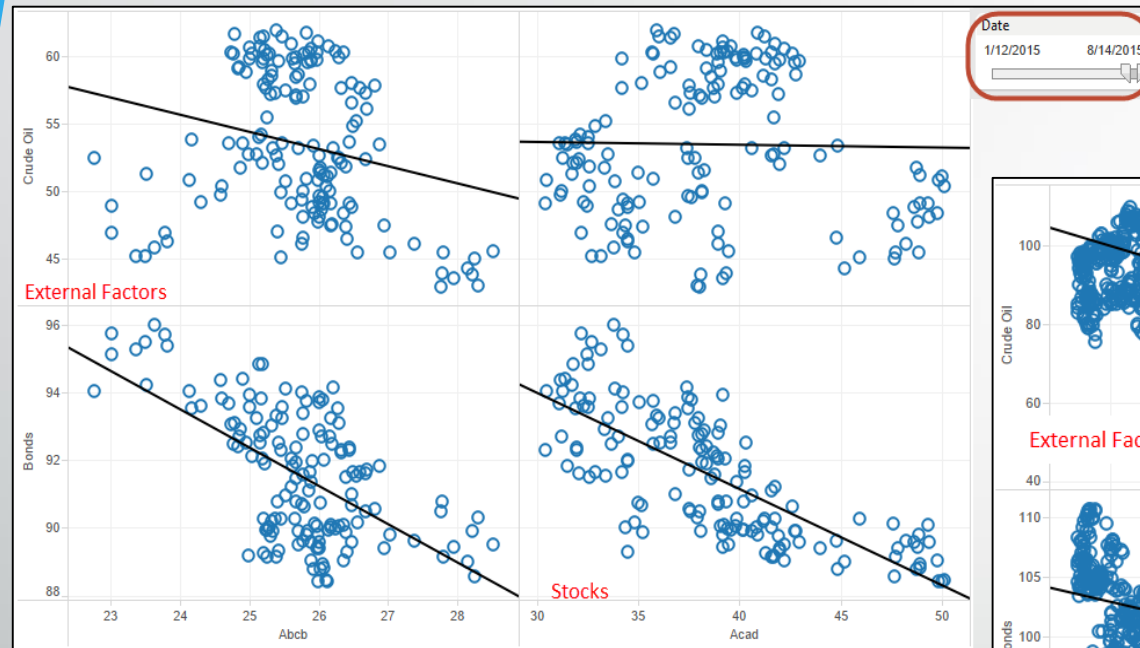
- Note: It was observed that there was no direct relationship between stocks and external factors as they vary across time

Zeppelin – Dynamic Form

Prediction – Impacted Stocks

Thanks !!!