Job Market Analysis System

Apoorva Ganapati - ag8159
Manan Chawda - mrc9419
Rohan Sardana - rs7445
Srijan Malhotra - sm9439

12/14/22

Introduction

Current Landscape

- With changing market conditions and looms of recession, job search becomes a complex process.
- The market volatility, workplace news and reviews on social media are becoming more relevant.
- There is a dearth of job market analytics tools to provide a detailed view of market sentiments for a company.

Problem Statement

We aim to solve the problem of job market analysis by developing a system that will leverage real-time data from multiple sources for streaming, modeling and visualization that will be used by candidates to make decisions on the job application or switch.

NYU

Big Data to the rescue!

With humongous real time data generated on social media and stocks, we can use

- Kibana:

  To deal with time series analysis of stock prices and custom visualizations for the social media data.

- PySpark:

  Ingest data into the system for employee reviews on Blind

- Elasticsearch:

  To store and analyse large volumes of different data points on stocks, tweets, reviews, that are being used for visualization.

Data Sources for the job market

An array of different data sources have been used for your project:

- Twitter:

  This data has been scraped using tweepy using Twitter Elevated Access Developer Account. It consist of Name, text, sentiment and follower count of the user

- Yahoo Finance:

  Live data is being pulled using the yahoo finance API. This gives us stock price data for the past 60 days, per min/day/week

- Blind Reviews:

  It has been used to analyze user reviews about different companies for their sentiment analysis..

NYU

Exploratory Data Analysis and Transformation

```
+-----------------+------------------+
|Review Sentiment |Number of Reviews |
+-----------------+------------------+
|        Positive |            32966 |
|         Neutral |             3412 |
|        Negative |             4032 |
+-----------------+------------------+
```

Fig 1. Sentiment Count of Blind reviews

WordCloud for reviews from Blind

Fig 2. WordCloud for review from Blind

Open	High	Low	Close	Volume	Dividends	Stock Splits

Fig 3. Real time stock price from Tweepy

Fig 4. Architecture Diagram

NYU

Architecture

Real time Streaming System:

1. **Data Ingestion** for live stock and twitter data
2. Subscription based **consumers**
3. **Spark** data analysis for reviews and historic data
4. **ElasticSearch** for storing ingested data
5. **Kibana** for visualization
6. Sentiment Analysis of Tweets using NLTK Sentiment Intensity Analyzer (**SIA**)
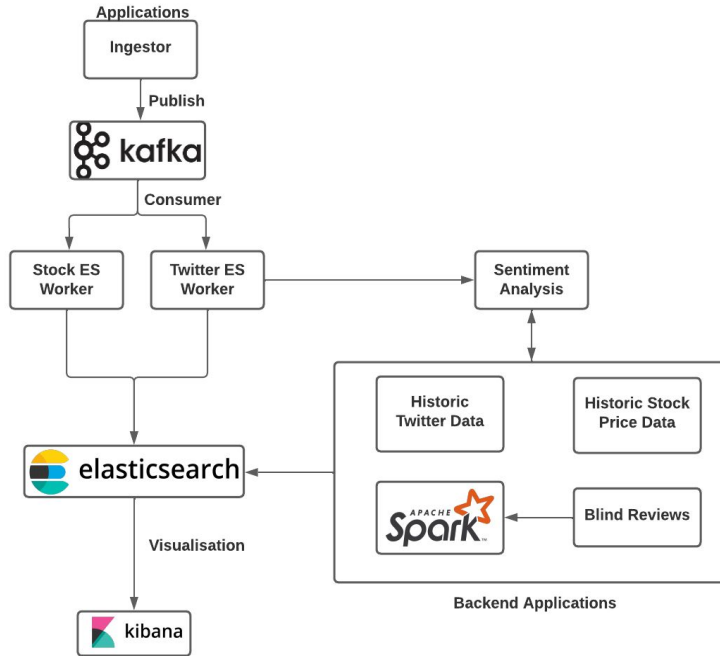
Visualisations

Fig 5. Kibana Dashboard view of Airbnb

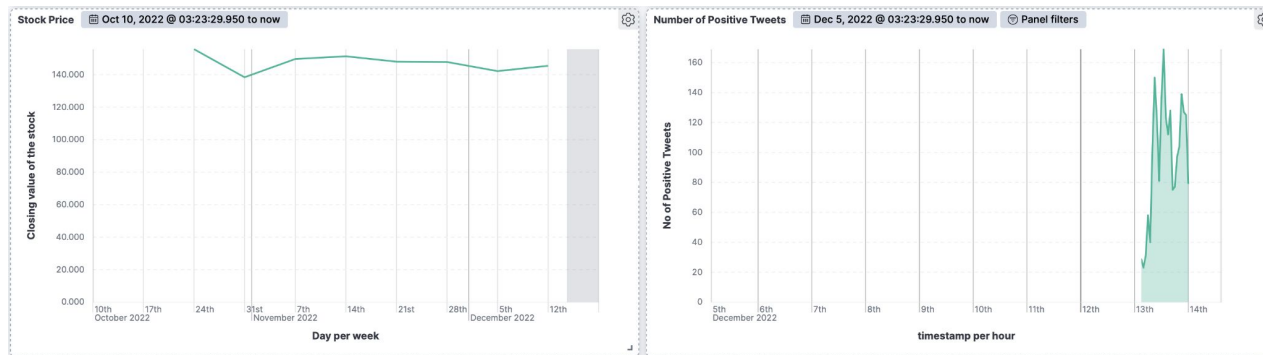Fig 6. Blind reviews WordCloud for Airbnb

NYU

7

Visualisations

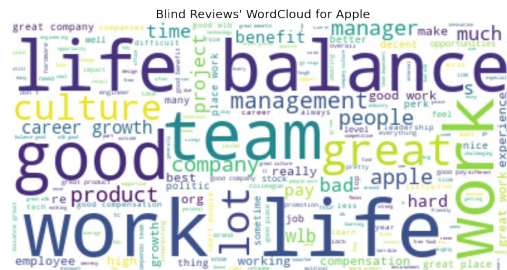Fig 7. Kibana Dashboard view of Apple

Fig 8. Blind reviews WordCloud for Apple

Demo

Job Market Analysis System

Future scope

- Launching dashboards as a chrome extension on job portals like Workday, Indeed, Handshake
- Using Premium Open API's of Glassdoor, LinkedIn and Twitter to ingest more employee reviews and market insights
- Using Ratings like S&P - Credit ratings for debts and soundness of a company.
- Private/non-listed companies data ingestion using news and funding information for performance tracking.

NYU

Thank You!

NYU