Individual Contribution Report

IRISH Stock Market Analysis

Student:

Bharat Jethwani

Teacher:

Dr. Shazia A. Afzal

Course:

Data Storage Solutions for Data Analytics

# Brainstorming

I was working for the last 3 years, so there was a continuous flow of money into my bank account, but when I used to do my books every quarter I observe hardly any savings in there. Also, the savings I had were in fixed deposits or Bonds, honestly, I never liked their returns. My friends suggested me to invest in shares, but I was always scared because of the stores, so never tried. And one fine day, when I was sitting with my friend Neeraj he mentioned a news about how Brexit is affecting the Irish stock marked. That is when I suggested analyzing the Irish Stock Market as a part of our Data Storage Solutions project. My aim was to remove all the factors which are affecting Irish share market and do a deep analysis on same. So that in future if an amateur has to invest he/she can look into these analysis reports and factors.

# Understanding Concepts

|  |  |  |
| --- | --- | --- |
| Because of curiosity and deep interest in python, I started learning web scrapping.  And started scraping Irish Stock details for ISE and all the news details related to those stocks and PESTLE data from Irish times. Well this was the first thing I learnt Web scrapping because of this project.  The thing about scraping data from multiple sources is that the data is very unclean. So, Neeraj had to clean all the scraped data using R. I and Aayush then split the cleaning work and finally we had the cleaned dataset.  Because of the same Neeraj and Aayush mentioned rvest to clean and sort data, so I started exploring in the same and learnt a few things about that.  The next step was to create ER model and we showed the same to our lecturer, because of her insights we made a few  had a lot many unmatched columns when I was adding tables directly, so I had to come up with SQL squery for the same. And add error outputs to flat file sources and do the analysis and update the query for the same.  The next step was to create reports using SSRS well visualization always gives me chills so I started with R where I used ggplot 2 to create various reports comparing different tables. Then to analyze the factors I wrote a wordcount code in R for all the Headlines.  . |  | Changes in our model. We went for Star Schema here.  Well I had been trained before in MySQL, so I had a prior knowledge about the concepts of Database, SQL queries etc. This module gave me a chance to use all those concepts in reality which was very exciting for me .  Well I had knowledge about databases but this was the first time I worked on ETL, right from the scratch, well I started loading data in our source tables and it wasn’t easy, I faced un countable errors for the same. Well few sleep less nights got me through those errors. Next step was to load data into Dimensions, So I modified our schema and started loading the data, well this was a but simple because of the previous errors. Next step was to load data into fact tables well this was interesting again because we  Aayush handled SSRS and taught me the same.  Helped Prasad in Neo4j because of priror knowledge on GPS Systems.  Also, I had the opportunity to learn about XML Schema and explored the concept of graph databases using Neo4j. |

# Implementation of Aforementioned Concepts in the Project

## Relational Databases

Contributed in creating the ERD model and the dimensional model based on the things I had learnt in class.

## Data Warehousing

I contributed in loading source tables and dimension tables in SSIS And individually created fact helped with developing reports in SSRS.

## Extensible Markup Language(XML)

Learnt from Neeraj about the same.

## Graph Databases (Neo4j)

Heleped Prasad for the same.Understood how to load CSV into Graph Databases.

# Challenges Faced

|  |  |  |
| --- | --- | --- |
| Getting Used To ETL and NEO4j |  | Cleaning of Data |
| Had a tough time initially with ETL faced many errors but learnt a lot troubleshooting the same. Neo4J was tough we faced few problems in relationships but got through the same using few different combinations. |  | One of the primary challenges I had to face initially was during cleaning of data. The scraped data was unclean with null/NA/garbage values. I had to ensure that the dataset was totally clean before using it for analysis. I did so by using R |
|  |  |  |
|  |  |  |
|  |  |  |

# Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit.

|  |  |
| --- | --- |
|  |  |