***Individual Contribution in the project-***

**“Predictive Movie Analytics”**

Student:

Rakshit Ratan

Student No:

10542705

Professor

Bernie Lydon

Course:

Data Storage Solutions for Data Analytics

**Context**

I looked up different databases as I have a keen interest to indulge in analysing data and studying specifics. After thorough digging, We as a group have chosen World Wide database from Relational Dataset Repository source and have done case study of the subject for this database.

**Initial step:**  
  
• Found a sample database dataset for Relational Dataset Repository

• Performed some changes according to the feedback of my group mates.

• Decided how to go with my group mates, and began to work on SQL.

**Explanatory material**

I learned about the primary keys and foreign keys in the tables and the relations between them, I also investigated source database tables to find the relationship between the tables and used these relationships to determine what data can be extracted, transformed and loaded to the destination database which is the data ware house. I created the Dimension tables using create query in the sql server, created constraints and Stored procedures for the same and Inserted the data in the dimension table using ADO net source and ADO net destination. I have learned the use of identity(1,1) in SQL while creating dimension tables in the data ware house. I also acquired knowledge about how to create a Star schema in the Database diagram and verify the mapping of primary key and foreign key in the Schema.

# SSMS, SSIS

I have learned how to use sort, merge, ADO net source, Ado net destination, Derived column and Lookup functions to Populate the data from source database to destination database. I Went back and forth between source tables and dimension tables to get accurate data for mapping into the destination fact table. After populating the data for all the dimension tables using SSIS, I have learned how to map the Ado net source to lookups to populate data in the destination table (fact table). I have created formulas to calculate different Tax Amount, Profit and Total Amount Including Tax and Excluding Tax through the derived column function which will finally calculate and will reach their fact table. I have learned about different SQL functions such as order, sort, distinct, UNION, COUNT, AVERAGE, MAX etc. I have learned many types of joins like inner, cross and full etc. to populate data from more than one table. Verified the data in the dimension and fact table using sql queries once data is populated in the dimension and fact table. I have gained knowledge on how to Extract, Transform and Load data to gain insights on the result produced by the data for example : Data obtained about Total transactions per city may be helpful to a company to understand in which cities their performance are best and which cities they can improve and grow.

**SSRS**

Created SSRS reports using stored procedure and build query, used data to prepare logical outcome such as transactions per city, Total Transactions City Wise, Transaction Amount Per Stock item Per Customer. Made Data base connection and create 5 reports, I have learned Group functions and Group properties while creating reports example: create a Parent Group for Customer Key and Visualised it through Hide and learned how to created parameterised report using SSRS tool. I have found different aims from fact table to create reports and write queries to design reports and using sum function calculated total profit per item.

**Tableau**

I learned how to generate graphs using Tableau and how the data is populated.

**Neo4j**

Learned how to match tables, create constraints and understand graphical relationships in the data, also learned how to import data to the neo4j browser from the relational database.

**Challenges Faced**

I was met by a few hurdles along the way but could overcome them efficiently by using online resources such as class room lectures, W3 schools, YouTube videos and Stack over flow.

**Conclusion**

* **Sales of Movie tickets depend on the Cast of Movie**
* **Movies Sales are high in particular period of an Year**
* **Movies with Lower Ticket Prices gets higher views**
* **We have used relational and dimensional modelling to design a data warehouse.**
* **We fetched the data and moved it across the data warehouse tables using ETL and we used SQL queries to insert data into the fact table and dimensions.**
* **We have used different types of visualization using SSRS package and Tableau.**
* **We used Neo4j to create graphical databases and created relations between them.**