SQL Project Report

* Dataset 1

It is a public data set of Chicago public school data which I got accessed through the course which I did by IBM named “Databases and SQL for Data Science with python”. It includes a variety of data related to the schools, students, teachers, and administration within the CPS system. This data can encompass a wide range of information, such as:

1. School performance metrics
2. Enrolment statistics
3. Demographics of students and staff
4. Attendance records
5. Budget and spending details
6. Standardized test scores
7. Graduation rates
8. Disciplinary actions

* Dataset 2

The Telecom Churn dataset contains information about customers of a telecommunications company, focusing on their demographics, service usage, and reasons for churn (discontinuing services). The key attributes in this dataset include:

* Customer ID: Unique identifier for each customer.
* Age: Age of the customer.
* City: City of residence.
* Zip Code: Residential zip code.
* Tenure in Months: Duration the customer has been with the company.
* Monthly Charge: Monthly charge for the customer's services.
* Total Charges: Total charges billed to the customer.
* Customer Status: Status of the customer (e.g., Active, Inactive, Churned).
* Churn Category: Category indicating the reason for churn.
* Churn Reason: Specific reason for customer churn.

This dataset is valuable for analyzing customer retention, identifying factors contributing to churn, and developing strategies to improve customer satisfaction and reduce churn rates.

* Learnings

I got to learn how to use IBM Db2, store data in it and apply queries on the table and download the output which we required. And also, how to use ibm\_db library in python to directly connect to the database and run query in the notebook.

Then I got to learn about SQLite database how to create your own serverless database and establish connection and use cursor to execute queries. And how to use magic SQL command in Jupiter notebook.

As the data needed some cleaning, So I cleaned it. And then I practiced as many queries as I can and try get maximum insights from the table.

The information I was able to extract from the dataset 1 are:

* Schools with the highest/lowest safety score
* Schools with the highest/lowest average attendance
* No. of student enrolled in the college as per community area
* Health certified school as per the area
* Schools with high teachers scores, environment score and much more

The information I was able to extract from the dataset 2 are:

* Number of customer by city
* Customers with High Average Monthly Long Distance Charges
* Total Revenue per City
* Total Revenue per City
* Churn Rate by Zip Code