DATA ANALYST

Working process of data analyst
1- Start with a problem statement.
2- Discover/Gather the data.
3- Cleanse the data.
4- Organize/transform the data.
5- Model the data.
6- Generate insights.

Project 1: Sales insights

Generating sales insights for a consumer goods company using SQL and Power BI.

Problem Statement:

AtliQ hardware company is a computer hardware & pheripheral manufacturer. They have many store across India. Sales director of this company interested in simple and efficient way of taking insights (" a picture is worth a thousand words"). He interested in dashboard (data driven decisions) that helps in increase the sales of a company.

Project Planning in Aimsgrid:

AIMS Grid

PURPOSE

To unlock sales insights that are not visible before for sales team for decision support & automate them to reduced manual time spent in data gathering.

STAKEHOLDERS

- Sales Director
- Marketing Team
- Customer Service Team
- Data & Analytics Team
- IT

END RESULT

An automated dashboard providing quick & latest sales insights in order to support data driven decision making.

SUCCESS CRITERIA

- Dashboard(s) uncovering sales order insights with latest data available
- Sales team able to take better decisions & prove 10 % cost savings of total spend
- Sales Analysts stop data gathering manually in order to save 20% of their business time and reinvest it value added activity

Discover/Gather data:

Mysql is also called OLTP (Online Transaction Processing System). It is very critical system, cannot afford that system to go down if that goes down then regular sales operations will get hampered.
If we directly perform analytical queries on mysql database then two problems will arise

1-Database go down then mainstream business is affected.
2-Data stored in OLTP is not often in a format that you want in order to perform analysis.

Because of these problems data from OLTP is extracted then transformation is done and load that data in to dataware house (e.g terradata, snowflake and so on). That transformed data is best for analytical queries. Transformation happens in ETL Tools (Extract, Transform and Load) e.g informatica or apache nifi. Data engineers do all these transformations and dataware house infrastructure. After that analyst pulls the data from dataware house and build powerBI dashboard. For simplicity, here directly plug powerBI to sql database.

Data Cleaning and Data Transformation:

Use Power Query (ETL tool) for data transformation.
Data cleaning for this project requires:
1-Removing duplicates from transaction table.
2-Convert the USD currency (b/c there are only two rows with USD currency other are INR currency).
3-Remove rows with missing values (how to handle with missing values based on the nature of the data and the analysis requirements. Here removing such rows is suitable).
4-Remove transactions with sales_amount equals to 0 or -1.

Data modelling

Use PowerBI for Data modelling.

Generate insights

Create Dashboard from that dashboard multiple insights can be generated or we can say that many questions can be answered.

Data Analysis Using SQL

1. Show all customer records

   `SELECT * FROM customers;`

1. Show total number of customers

   `SELECT count(*) FROM customers;`

1. Show transactions for Chennai market (market code for chennai is Mark001

   `SELECT * FROM transactions where market_code='Mark001';`

1. Show distrinct product codes that were sold in chennai

   `SELECT distinct product_code FROM transactions where market_code='Mark001';`

1. Show transactions where currency is US dollars

   `SELECT * from transactions where currency="USD"'`

1. Show transactions in 2020 join by date table

   `SELECT transactions.*, date.* FROM transactions INNER JOIN date ON transactions.order_date=date.date where date.year=2020;`

1. Show total revenue in year 2020,

   `SELECT SUM(transactions.sales_amount) FROM transactions INNER JOIN date ON transactions.order_date=date.date where date.year=2020 and transactions.currency="INR\r" or transactions.currency="USD\r";`

1. Show total revenue in year 2020, January Month,

   `SELECT SUM(transactions.sales_amount) FROM transactions INNER JOIN date ON transactions.order_date=date.date where date.year=2020 and and date.month_name="January" and (transactions.currency="INR\r" or transactions.currency="USD\r");`

1. Show total revenue in year 2020 in Chennai

   `SELECT SUM(transactions.sales_amount) FROM transactions INNER JOIN date ON transactions.order_date=date.date where date.year=2020
and transactions.market_code="Mark001";`

Data Analysis Using Power BI

Data Analysis Expressions (DAX) is a library of functions and operators that can be combined to build formulas and expressions in Power BI, Analysis Services, and Power Pivot in Excel data models.

Some DAX used in this project are:

1-Use of DAX in data cleaning
Formula to create norm_amount column

`= Table.AddColumn(#"Filtered Rows", "norm_amount", each if [currency] = "USD" or [currency] ="USD#(cr)" then [sales_amount]*75 else [sales_amount], type any)`

2-Use of DAX in creating measures

Revenue = SUM('transaction new'[norm_currency])

Total Profit Margin = SUM('transaction new'[profit_margin])

sales Qty = SUM('transaction new'[sales_qty])

Revenue LY = CALCULATE([Revenue],SAMEPERIODLASTYEAR('date'[date]))

RevenueContribution=DIVIDE([Revenue],CALCULATE([Revenue],ALL('products'),ALL('customers'),ALL('market')))

Profit Margin Contribution% = DIVIDE([Total Profit Margin],CALCULATE([Total Profit Margin],ALL('products'),ALL('customers'),ALL('market')))

profit Margin % = DIVIDE([Total Profit Margin],[Revenue])

Using these measures to create visuals and taking insights from these visuals