AMAZON SALES ANALYSIS

SQL-PROJECT

Linkdin (in)

Github

Purposes Of The Project :

THE MAJOR AIM OF THIS PROJECT IS TO GAIN INSIGHT INTO THE SALES DATA OF AMAZON TO UNDERSTAND THE DIFFERENT FACTORS THAT AFFECT SALES OF THE DIFFERENT BRANCHES.

Github

ANALYSIS LIST:

1. PRODUCT ANALYSIS

- Conduct analysis on the data to understand the different product lines, the products

2. SALES ANALYSIS

- This analysis aims to answer the question of the sales trends of product. The result of this can help us measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

3. CUSTOMER ANALYSIS

- This analysis aims to uncover the different customer segments, purchase trends and the profitability of each customer segment.

Github

APPROACH USED:

1. DATA WRANGLING:

- This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace missing or NULL values.

- 1.1   Build a database

- 1.2   Create a table and insert the data.

- 1.3   Select columns with null values in them. There are no null values in our database as in creating the tables, we set NOT NULL for each field, hence null values are filtered out.

Github

APPROACH USED:

2. FEATURE ENGINEERING:

- This will help us generate some new columns from existing ones.

- 2.1   Add a new column named **timeofday** to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.

- 2.2   Add a new column named **dayname** that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.

- 2.3   Add a new column named **monthname** that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.

Github

APPROACH USED:

3.EXPLORATORY DATA ANALYSIS (EDA):

- Exploratory data analysis is done to answer the following questions and aims of this project.

Github

FEATURE ENGINEERING: (ADD 'MONTHNAME')

- **Step 1: Add a new column 'monthname'.**

  -ALTER TABLE amazon ADD COLUMN MonthName VARCHAR(20);

- **Step 2: Update the new column with the 'monthname'.**

```sql
UPDATE amazon SET monthname = CASE
        WHEN MONTH(Date) = 1 THEN 'January'
        WHEN MONTH(Date) = 2 THEN 'February'
        WHEN MONTH(Date) = 3 THEN 'March'
        WHEN MONTH(Date) = 4 THEN 'April'
        WHEN MONTH(Date) = 5 THEN 'May'
        WHEN MONTH(Date) = 6 THEN 'June'
        WHEN MONTH(Date) = 7 THEN 'July'
        WHEN MONTH(Date) = 8 THEN 'August'
        WHEN MONTH(Date) = 9 THEN 'September'
        WHEN MONTH(Date) = 10 THEN 'October'
        WHEN MONTH(Date) = 11 THEN 'November'
ELSE 'December'
END;
```

Github

- **Step 1: Add a new column 'Days'.**

  -ALTER TABLE amazon ADD COLUMN Days VARCHAR(20);

- **Step 2: Update the new column with the 'Days' name.**

```
SET Days =
    CASE dayofweek(date)
        WHEN 1 THEN 'Sunday'
        WHEN 2 THEN 'Monday'
        WHEN 3 THEN 'Tuesday'
        WHEN 4 THEN 'Wednesday'
        WHEN 5 THEN 'Thursday'
        WHEN 6 THEN 'Friday'
        WHEN 7 THEN 'Saturday'
    END;
```

Github

FEATURE ENGINEERING: (ADD 'TIMEOFDAY')

- **Step 1: Add a new column 'TimeOfDay'.**

    -ALTER TABLE amazon ADD TimeOfDay VARCHAR(20);

- **Step 2: Update the new column with the 'TimeOfDay' name.**

```sql
UPDATE amazon
SET TimeOfDay = CASE
    -- Morning: From 00:00:00 to 11:59:59
    WHEN EXTRACT(HOUR FROM Time) >= 0 AND EXTRACT(HOUR FROM Time) < 12 THEN 'Morning'
    -- Afternoon: From 12:00:00 to 17:59:59
    WHEN EXTRACT(HOUR FROM Time) >= 12 AND EXTRACT(HOUR FROM Time) < 18 THEN 'Afternoon'
    -- Evening: From 18:00:00 to 23:59:59
    ELSE 'Evening'
END;
```

Github

QUESTION NO.1

Q1 : What is the count of distinct cities in the dataset?

```sql
select Count(Distinct city) as 'No of Distinct Cities'
from amazon;
```
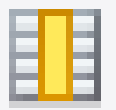
IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **The DISTINCT statement is used to return only distinct (different) values.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name('No of Distinct Cities').**

Result Grid	Filter Rows:
No of Distinct Cities	
▶ 3	

Github

QUESTION NO.2

Q2 : For each Branch, what is the corresponding city?

```
select

Distinct Branch,City

from amazon ORDER BY Branch;
```

IN THIS,

- **SELECT statement is used to select data from a database.**
- **The DISTINCT statement is used to return only distinct (different) values.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Result Grid		Filter
	Branch	City
▶	A	Yangon
	B	Mandalay
	C	Naypyitaw

Github

QUESTION NO.3

Q3 : What is the Count of Distinct product lines in the dataset?

```sql
select 
Count(distinct `Product line`) as 'No. Of Product line'
from amazon;
```

No. Of Product line
6

Result Grid | Filter Row

IN THIS,

- **SELECT** statement is used to select data from a database.
- **DISTINCT** statement is used to return only distinct (different) values.
- The **COUNT()** function returns the number of rows that matches a specified criterion.
- **ALIASES** (as) are used to give a table, or a column in a table, a temporary name('No. Of Product line').

Github

QUESTION NO.4

Q4 : Which payment method occurs most frequently?

```sql
select Payment,
count(*) as Frequently_Payment_Method
from amazon
group by Payment
order by Frequently_Payment_Method desc;
```

Result Grid | Filter Rows:

Payment	Frequently_Payment_Method
Ewallet	345
Cash	344
Credit card	311

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.5

Q5 : Which product line has the highest sales?

```sql
select `Product line`,
Count(`Invoice ID`) as `Sales Count`
from amazon
group by `Product line`
Order by `Sales Count` desc limit 1;
```

	Product line	Sales Count
▶	Fashion accessories	178

Result Grid | Filter Rows:

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**
- **The LIMIT clause is used to specify the number of records to return.**

Github

QUESTION NO.6

Q6 : How much revenue is generated each month?

```sql
select Monthname,
count(total) as Revenue
from amazon
group by Monthname
order by Revenue desc;
```

Result Grid | Filter Rows:

	Monthname	Revenue
▶	January	352
	March	345
	February	303

IN THIS,

- The SELECT statement is used to select data from a database.
- The COUNT() function returns the number of rows that matches a specified criterion.
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

QUESTION NO.7

Q7 : In which month did the cost of goods sold reach its peak?

```sql
select Monthname,
sum(cogs) as cogs_sold
from amazon group by Monthname
order by cogs_sold desc;
```

	Monthname	cogs_sold
▶	January	110754.16000000002
	March	104243.33999999997
	February	92589.88

Result Grid | Filter Rows:

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The SUM() function returns the total sum of a numeric column.**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.8

Q8 : Which product line generated the highest revenue?

```sql
select `Product line`,
Sum(Total) as Total_Revenue
from amazon group by `Product line`
order by Total_Revenue desc limit 1;
```

Result Grid	Filter Rows:	
	Product line	Total_Revenue
▶	Food and beverages	56144.84400000005

IN THIS,

- The SELECT statement is used to select data from a database.
- The SUM() function returns the total sum of a numeric column.
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.
- The LIMIT clause is used to specify the number of records to return.

Github

QUESTION NO.9

Q9 : In which city was the highest revenue recorded?

```sql
select distinct City ,
sum(Total) as Revenue
from amazon
group by City
order by Revenue desc;
```

Result Grid		Filter Rows:	

	City	Revenue
▶	Naypyitaw	110568.70649999994
	Yangon	106200.3705000001
	Mandalay	106197.67199999996

IN THIS,

- The SELECT statement is used to select data from a database.
- The SUM() function returns the total sum of a numeric column.
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

QUESTION NO.10

Q10 : Which product line incurred the highest Value Added Tax?

```
select distinct City ,
sum(Total) as Revenue
from amazon
group by City
order by Revenue desc;
```

Result Grid		Filter Rows:	
	City	Revenue	
▶	Naypyitaw	110568.70649999994	
	Yangon	106200.3705000001	
	Mandalay	106197.67199999996	

IN THIS,

- The SELECT statement is used to select data from a database.
- The SUM() function returns the total sum of a numeric column.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name.
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

Q11 : For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."

```sql
select `Product line`,sum(Total) as TotalSales,
case when
sum(Total) > (select Avg(Total) as avg from amazon) then "Good"
else "Bad" end as SalesType
from amazon group by `Product line`
order by TotalSales desc;
```

	Product line	TotalSales	SalesType
▶	Food and beverages	56144.844000000005	Good
	Sports and travel	55122.826499999996	Good
	Electronic accessories	54337.531500000005	Good
	Fashion accessories	54305.895	Good
	Home and lifestyle	53861.91300000001	Good
	Health and beauty	49193.739000000016	Good

IN THIS,
- The SELECT statement is used to select data from a database.
- The SUM() function returns the total sum of a numeric column.
- The CASE statement goes through conditions and returns a value when the first condition is met (like an if-then-else statement). So, once a condition is true, it will stop reading and return the result. If no conditions are true, it returns the value in the ELSE clause.If there is no ELSE part and no conditions are true, it returns NULL.
- In this we used NON-CORREALATED SUBQUERY.The inner query doesn't depend on the outer query and can run as a stand-alone query
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.

Github

QUESTION NO.12

Q12 : Identify the branch that exceeded the average number of products sold.

```sql
select Distinct Branch,
Sum(Quantity) as TOTALQuantity
from amazon group by Branch
having TOTALQuantity > (select AVG(Quantity) from amazon) order by branch;
```

Result Grid | Filter Rows:

Branch	TOTALQuantity
A	1859
B	1820
C	1831

IN THIS,

- The SELECT statement is used to select data from a database.
- The SUM() function returns the total sum of a numeric column.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name(TOTALQuantity).
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The HAVING clause was added to SQL because the WHERE keyword cannot be used with aggregate functions.
- In this we used NON-CORREALATED SUBQUERY.The inner query doesn't depend on the outer query and can run as a stand-alone query
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

QUESTION NO.13

Q13 : Which product line is most frequently associated with each gender?

```sql
select `Product line`,Gender,
Count('Invoice ID') as Frequency
from Amazon group by `Product line`,Gender
order by Count('Invoice ID') desc;
```

Result Grid	Filter Rows:		
Product line	Gender	Frequency
Fashion accessories	Female	96
Food and beverages	Female	90
Health and beauty	Male	88
Sports and travel	Female	88
Electronic accessories	Male	86

IN THIS,

- The SELECT statement is used to select data from a database.
- The COUNT() function returns the number of rows that matches a specified criterion.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name(Frequency).
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

QUESTION NO.14

Q14 : Calculate the average rating for each product line.

```sql
select `Product line`,
Avg(Rating) as AVG_RATING
from Amazon group by `Product line`
order by AVG_RATING desc;
```

Product line	AVG_RATING
Food and beverages	7.113218390804598
Fashion accessories	7.029213483146067
Health and beauty	7.003289473684212
Electronic accessories	6.92470588235294
Sports and travel	6.91626506024964

IN THIS,

- The SELECT statement is used to select data from a database.
- The AVG() function returns the average value of a numeric column.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name(AVG_RATING).
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

QUESTION NO.15

Q15 : Count the sales occurrences for each time of day on every weekday.

```sql
select Days,TimeOfDay,
Count('Invoice ID') as 'Sales Count' from amazon
where Days not in ('Saturday','Sunday')
group by Days,TimeOfDay order by Days,'Sales Count';
```

Days	TimeOfDay	Sales Count
Friday	Afternoon	74
Friday	Evening	36
Friday	Morning	29
Monday	Afternoon	75
Monday	Evening	29
Monday	Morning	21
Thursday	Afternoon	76

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **The WHERE clause is used to filter records.It is used to extract only those records that fulfill a specified condition.**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.16

Q16 : Identify the customer type contributing the highest revenue.

```sql
select `Customer type`,
Sum(Total) as 'Highest Revenue'
from amazon group by `Customer type`
order by 'Highest Revenue';
```

Result Grid	Filter Rows:	
Customer type	**Highest Revenue**
▶ Member	164223.44400000002
Normal	158743.30500000005

IN THIS,
- **The SELECT statement is used to select data from a database.**
- **The SUM() function returns the total sum of a numeric column.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name('Highest Revenue').**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.17

Q17 : Determine the city with the highest VAT percentage.

```sql
select City ,
sum(`Tax 5%`) as 'Highest VAT%'
from amazon group by City
order by 'Highest VAT%' ;
```

	City	Highest VAT%
▶	Yangon	5057.160500000002
	Naypyitaw	5265.176500000002
	Mandalay	5057.032000000003

Result Grid | Filter Rows:

IN THIS,

- The SELECT statement is used to select data from a database .
- The SUM() function returns the total sum of a numeric column.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name('Highest VAT%').
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

Q18 : Identify the customer type with the highest VAT payments.

```
select `Customer type`,
sum(`Tax 5%`) as 'Highest VAT%'
from amazon group by `Customer type`
order by 'Highest VAT%';
```

	Customer type	Highest VAT%
▶	Member	7820.164000000002
	Normal	7559.205000000003

Result Grid | Filter Rows:

IN THIS,
- The SELECT statement is used to select data from a database AND
- The SUM() function returns the total sum of a numeric column.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name('Highest VAT%').
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

Q19 : What is the count of distinct customer types in the dataset?

```
select count(distinct `Customer type`) as `Customer type`
from amazon;
```

Result Grid
Customer type
▶ 2

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **The DISTINCT statement is used to return only distinct (different) values.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name('`Customer type`).**

Github

Q20 : What is the count of distinct payment methods in the dataset?

```sql
select Count(distinct Payment) as Payment_Method
from amazon;
```

Result Grid		Fi
	Payment_Method	
▶	3	

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **The DISTINCT statement is used to return only distinct (different) values.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name(Payment_Method).**

Github

QUESTION NO.21

Q21 : Which customer type occurs most frequently?

```sql
select `Customer type`,
Count(`Invoice ID`) as 'No. Of Frequency'
from amazon group by `Customer type`;
```

	Customer type	No. Of Frequency
▶	Member	501
	Normal	499

Result Grid | Filter Rows:

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name('No. Of Frequency').**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**

Github

Q22 : Identify the customer type with the highest purchase frequency.

```sql
select `Customer type`,
Count(`Invoice ID`) as 'Highest Purchase Frequency'
from amazon group by `Customer type`;
```

Result Grid	Filter Rows:	
	Customer type	Highest Purchase Frequency
▶	Member	501
	Normal	499

IN THIS,
- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name("Highest Purchase Frequency").**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
-

Github

Q23 : Determine the predominant gender among customers.

```sql
select Gender,
Count(Gender) as Predominat_Gender
from amazon group by Gender
order by Predominat_Gender desc;
```

	Gender	Predominat_Gender
▶	Female	501
	Male	499

Result Grid | Filter Rows:

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name(Predominat_Gender).**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.24

Q24 : Examine the distribution of genders within each branch.

```sql
select Branch,Gender,
count(*) as Frequency from amazon
group by Branch, Gender
order by Branch,Frequency desc;
```

	Branch	Gender	Frequency
▶	A	Male	179
	A	Female	161
	B	Male	170
	B	Female	162
	C	Female	178
	C	Male	150

Result Grid | Filter Rows:

IN THIS,

- The SELECT statement is used to select data from a database.
- The COUNT() function returns the number of rows that matches a specified criterion.
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

QUESTION NO.25

Q25 : Identify the time of day when customers provide the most ratings.

```sql
select TimeOfDay,Count(Rating) as Most_Rating
from amazon group by TimeOfDay
order by Most_Rating desc;
```

	TimeOfDay	Most_Rating
▶	Afternoon	528
	Evening	281
	Morning	191

Result Grid | Filter Rows:

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The COUNT() function returns the number of rows that matches a specified criterion.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name(Most_Rating).**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.26

Q26 : Determine the time of day with the highest customer ratings for each branch.

```sql
select Branch,TimeOfDay,
AVG(Rating) as AVG_Rating
from amazon group by Branch,TimeOfDay
order by Branch,AVG_Rating desc;
```

Branch	TimeOfDay	AVG_Rating
A	Afternoon	7.0567567567567595
A	Morning	7.005479452054794
A	Evening	6.979268292682928
B	Morning	6.891525423728813
B	Afternoon	6.806790123456792
B	Evening	6.795495495495495
C	Afternoon	7.0955801104972345

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The AVG() function returns the average value of a numeric column.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name(AVG_Rating).**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order..**

Github

Q27 : Identify the day of the week with the highest average ratings.

```sql
select Days,Avg(Rating) as Highest_AVG_Rating
from amazon group by Days
order by Highest_AVG_Rating desc;
```

	Days	Highest_AVG_Rating
▶	Monday	7.15359999999999
	Friday	7.076258992805756
	Sunday	7.011278195488723
	Tuesday	7.003164556962025
	Saturday	6.90182926829268
	Thursday	6.88985507246377
	Wednesday	6.805594405594405

IN THIS,

- **The SELECT statement is used to select data from a database.**
- **The AVG() function returns the average value of a numeric column.**
- **ALIASES (as) are used to give a table, or a column in a table, a temporary name(Highest_AVG_Rating).**
- **The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.**
- **The ORDER BY keyword is used to sort the result-set in ascending or descending order.**

Github

QUESTION NO.28

Q28 : Determine the day of the week with the highest average ratings for each branch.

```sql
select Days,Branch,
Avg(Rating) as Highest_AVG_Rating
from amazon group by Days,Branch
order by Highest_AVG_Rating desc;
```

	Days	Branch	Highest_AVG_Rating
▶	Monday	B	7.335897435897434
	Friday	A	7.311999999999985
	Friday	C	7.278947368421051
	Saturday	C	7.229629629629631
	Monday	A	7.097916666666666
	Sunday	A	7.078846153846157
	Wednesday	C	7.064000000000004

IN THIS,

- The SELECT statement is used to select data from a database.
- The AVG() function returns the average value of a numeric column.
- ALIASES (as) are used to give a table, or a column in a table, a temporary name(Highest_AVG_Rating).
- The GROUP BY statement groups rows that have the same values into summary rows. GROUP BY statement is often used with aggregate functions (COUNT(), MAX(), MIN(), SUM(), AVG()) to group the result-set by one or more columns.
- The ORDER BY keyword is used to sort the result-set in ascending or descending order.

Github

PRODUCT ANALYSIS

1.

There is 6 Product line given in Amazon Database.

2.

Most Attracting Product line for Female is Fassion Accessories.

3.

FOOD & BEVERAGES Product line has a HIGHEST REVENUE.

Github

SALES ANALYSIS

1.

NAPYTAW City has a Highest Revenue

2.

Highest Revenue Month is JANUARY.

3.

Member Cutomer has Contributed more for the Revenue.

Github

CUSTOMER ANALYSIS

1.

EWallet is the Most Frequently Payment Method Used by Customer.

2.

Predominant Customer is Female.

3.

Customer provide the most rating at afternoon time.

Github

RECOMENDATIONS & IMPROVEMENTS :

1. Underperformed Product line such as Health and Beauty could be improved by performing Offers for Customers.

2. Develop More Market strategies/Offers for loyal and other Customer segments.

3. Focus on product lines and customer segments generating the highest revenue and VAT.

4. Utilize customer ratings to improve product offerings and customer service. Focus on times of day and branches with lower average ratings to enhance the overall customer experience.

Github

ACHIEVMENTS:

- Gained comprehensive insights into sales data, enabling data-driven decision-making.
- Improved understanding of customer behavior and sales performance metrics.

Github

By implementing these suggestions based on the SQL analysis, Amazon can make data-driven decisions to optimize sales, enhance customer satisfaction, and improve overall business performance.

Thank you!

vinayrajbhar14@gmail.com

Github