**PORTFOLIO PROJECT**

**Professional Background**

My name is Duggani Rohith. I am from a non-IT background. I am a BSc. Agriculture graduate. Though, I’m a BSc graduate I learnt data analytics because it has grabbed by attention and I’ve learnt with interest and passion. I have completed 8 projects in SQL, Excel, for data analytics. These projects showcase my expertise in data analysis and interpretation. With strong analytical skills and proficiency in SQL, Excel, Tableau, Python. I bring valuable insights to the field of data analytics.

Most probably I am a work enthusiast. I prefer work first over other things. My skills are I am a fast learner, dedicated, I can communicate with people and mingle with them at ease.My IT skills are Python,MYSQL,Tableau ,Statistics, MS-Excel,and other Microsoft applications.

**Table Of Contents**

Professional Background--------------------------------------------------- 1

Table of Contents ----------------------------------------------------------- 2

Data Analytics process using real time example ----------------------------------- 3 -4

Instagram User Analytics Description------------------------------------ 4

Design ----------------------------------------------------------------------- -7

Findings ---------------------------------------------------------------------7-9

Data Analysis -------------------------------------------------------------9

Conclusion ------------------------------------------------------------------ 9

Operation and metric Analysis Description --------------------------------------9

The Problem -----------------------------------------------------------------10

Findings ----------------------------------------------------------------- 10-18

Conclusion -------------------------------------------------------------------19

Hiring process analytics------------------------------------------------------- 19

The Problem ----------------------------------------------------------------

Design ----------------------------------------------------------------------- 19-24

Findings ------------------------------------------------------------------

Conclusion---------------------------------------------------------------25

Impact of car features-------------------------------------------------25

Design ------------------------------------------------------------------------

Findings -------------------------------------------------------------------25-34

Conclusion --------------------------------------------------------------------34

Appendix----------------------------------------------------------------------34

**Project 1-Real- life situation  : A Fertilizer company improving crop yield  using Data Analytics**

**Plan:**A fertilizer company wants to improve their product offerings by analyzing their customers' data and market trends. The company aims to provide the best-suited fertilizers for various crops, increase their sales and enhance customer satisfaction.

**Prepare:**The fertilizer company collects data from farmers, including their location, crop type, soil type, and weather conditions. The company also gathers data on sales, product performance, and customer feedback. Data is collected from various sources, including surveys, social media, and market reports.

**Process:**The data collected from various sources is organized and processed for analysis. The company uses data cleaning, transformation, and feature engineering techniques to prepare the data for analysis. The processed data is stored in a data warehouse, where it can be accessed easily for analysis.

**Analyze:**The fertilizer company uses data analytics techniques like statistical analysis, data visualization, and machine learning to uncover patterns and insights in the data. The company analyzes the data to determine the most popular crop types, the types of soil and weather conditions where specific fertilizers are most effective, and how different products perform in the market. By analyzing the data, the company can identify areas for improvement and develop new products that meet the needs of their customers.

**Share:**The insights gained from the data analysis are shared with different departments, including the product development team and the marketing team. The product development team can use the data to develop new fertilizers that are tailored to specific crops and soil types. The marketing team can use the insights to create targeted marketing campaigns that focus on the benefits of the company's products.​

​

**Act:**The fertilizer company takes action based on the insights gained from the data analysis. The company creates new products that are tailored to specific crops and soil types based on the data analysis. The company also creates targeted marketing campaigns that focus on the benefits of their products. The company continues to collect data and analyze it to improve their product offerings and meet the needs of their customers.

**Conclusion**

In conclusion, data analytics is a powerful tool that can help a fertilizer company to improve their product offerings, increase sales, and enhance customer satisfaction. The data analytics process involves planning, preparing, processing, analyzing, sharing, and acting on data to gain insights that can drive business decisions. By analyzing data, the company can identify areas for improvement and develop new products that meet the needs of their customers.

2. Instagram User Analytics

**Project description**: This project is about Instagram user analytics which is to understand how users effectively use instagram and data collected from them based on their activities such as liking photos, following ,comments under photos,usage of hastags while uploading photos etc and based on that data we can make users to spend more time on instagram app by personalized algorithm,ads based on their likes  and improving our business.

**Approach:** After going through requirements trying to understand what information i wanted to make report to marketing ,production and development teams.Going through each tables and column and linking common columns associated between tables which gave an outlook of what all the datas given and retrieve so many outcomes to understand about users and their activities

**Tech-Stack Used**: I have used MY SQL workbench version 8.0 CE, to create database and run queries and db-fiddle.com

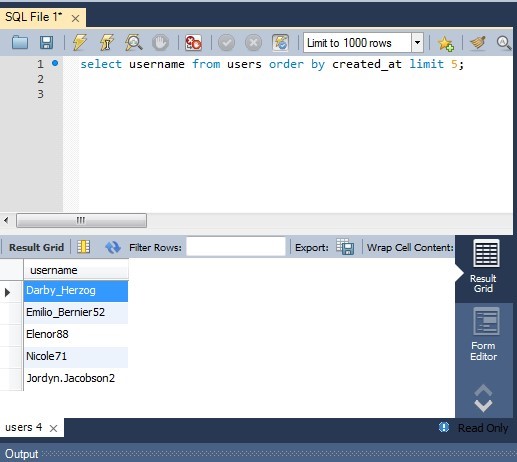
**Insights**: I have learned how to use join functions ,how and where to use sorting,aggregate functions ,so in declaring contest  winner problem using problem solution video I have learnt how to use more than one joins in a query,learnt to use day function on query,group by function and sorting functions.

**Result:** After  SQL learning concepts and practice problems,I could able to apply all concepts asked to solve and able to implement the outcome required for various team which will be helpful in improving business .

1.**Rewarding Most Loyal Users**

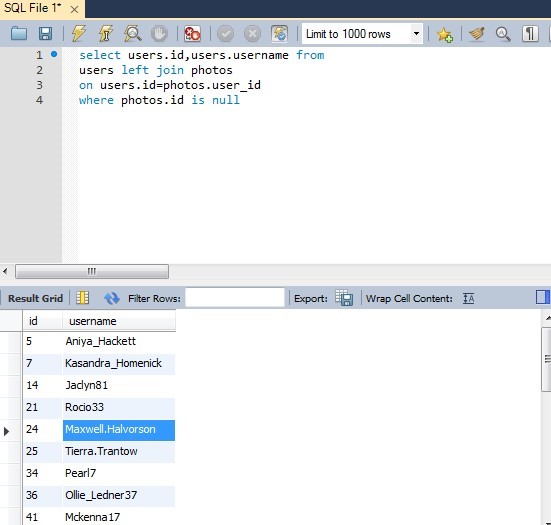
Your Task: Find the 5 oldest users of the Instagram from the database provided

**Query and Result:**



1. **Remind Inactive Users to Start Posting**

Your Task: Find the users who have never posted a single photo on Instagram **Query and Result:**



Output

5 Aniya\_Hackett

7 Kasandra\_Homenick

14 Jaclyn81

21 Rocio33

1. Maxwell.Halvorson
2. Tierra.Trantow

34 Pearl7

36 Ollie\_Ledner37

41 Mckenna17

45 David.Osinski47

49 Morgan.Kassulke

1. Linnea59
2. Duane60

57 Julien\_Schmidt

66 Mike.Auer39

68 Franco\_Keebler64 71 Nia\_Haag

1. Hulda.Macejkovic
2. Leslie67
3. Janelle.Nikolaus81
4. Darby\_Herzog
5. Esther.Zulauf61

83 Bartholome.Bernhard

1. Jessyca\_West
2. Esmeralda.Mraz57
3. Bethany20
4. **Declaring Contest Winner**

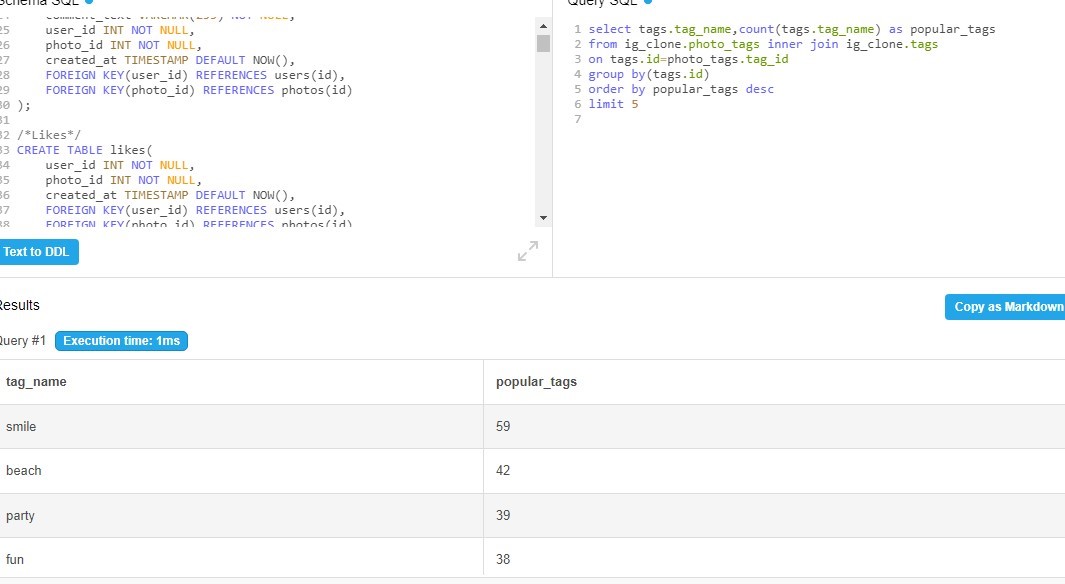
Your Task: Identify the winner of the contest and provide their details to the team

**Result:**

1.   
   **Hashtag Researching**

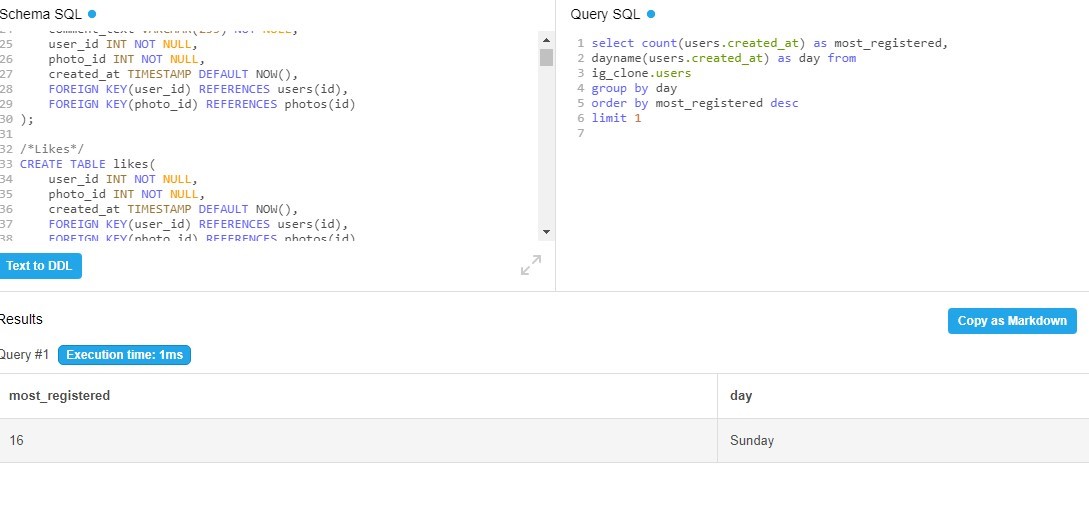
Your Task: Identify and suggest the top 5 most commonly used hashtags on the platform.

**Result:**

****

1. **Launch AD Campaign**

Your Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign:Based on results Sunday can be scheduled for ad campaign.



**B) Investor Metrics:** Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

1. **User Engagement:**

Your Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users **Query:**

select users.id,sum(users.id)as total\_users,sum(photos.id) as total\_photos, avg(photos.id)as posts from ig\_clone.users join ig\_clone.photos on users.id=photos.user\_id group by(photos.id) order by total\_users

**Result:**



2.**Bots & Fake Accounts:**

Your Task: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

**Query and Result:**

select users.username,count(likes.photo\_id) as bot from ig\_clone.photos join ig\_clone.likes on photos.id=likes.photo\_id join ig\_clone.users on users.id=likes.user\_id group by users.username order by bot desc

**3.Operation and metric analysis**

This project(OPERATION & METRIC ANALYTICS) is all about analysis for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. This kind of analysis is further used to predict the overall growth or decline of a company’s fortune.

**Approach:** We are going to create a Database in MySQL workbench and use MySQL queries to find out the required data from the data stored in the Database.

**Tech-Stack Used:** MySQL Workbench 8.0 CE,        to run MySQL Queries.

**Requirement:** We are going to find some Insights such as “NO OF JOBS REVIEWED”, “THROUGHPUT”, “PERCENTAGE SHARE OF EACH LANGUAGE”, “DUPLICATE ROWS”, “USER ENGAGEMENT”, and some metrices like “USER GROWTH”, “WEEKLY RETENTION”, “WEEKLY ENGANGEMENT”, “EMAIL ENGAGEMEMT”.

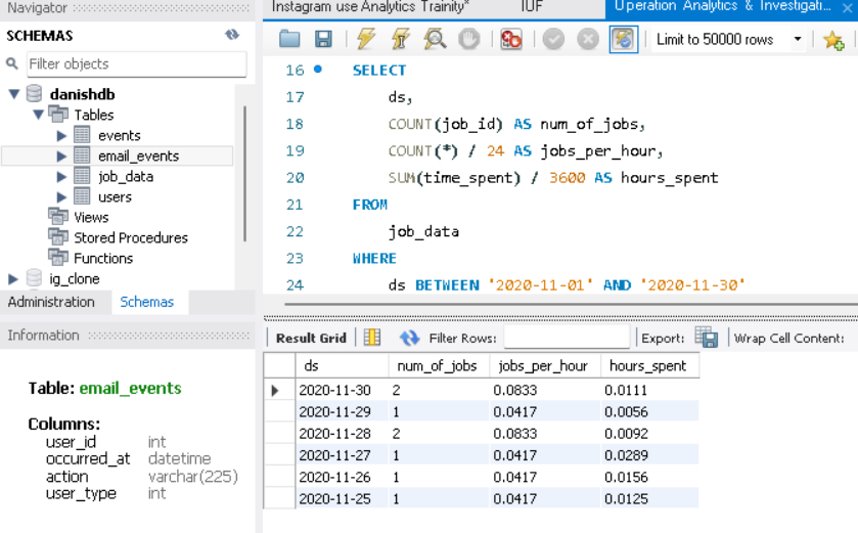
**NUMBER OF JOBS REVIEWED:**​

​

We have to calculate the  number of jobs reviewed per hour per day for the last 30 days’ .​

A simple query in MySQL will display the result as shown in the picture

**MySQL Query & Result:**



**THROUGHPUT:**​

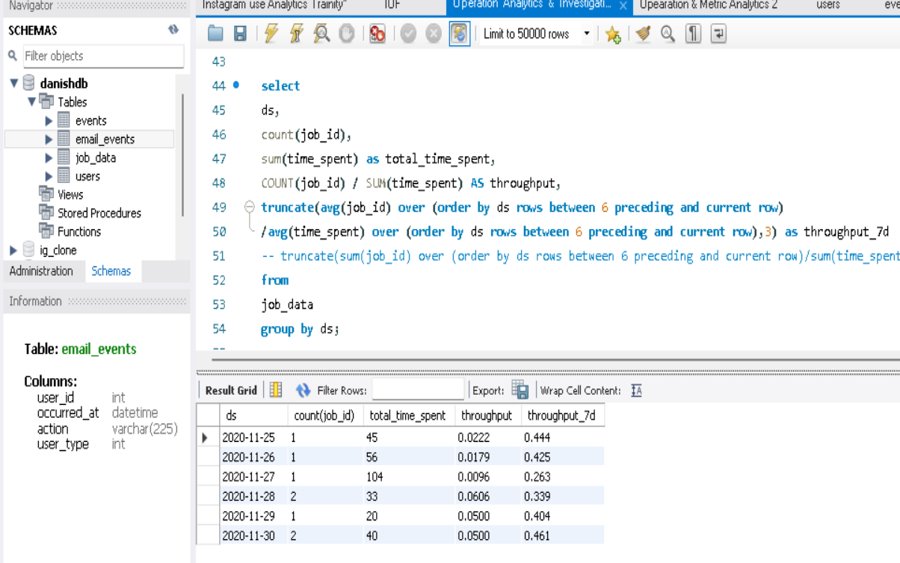
​

We have to find ‘No of events happening per second’ and ‘7 day Rolling Average and compare both to find the better Indicator’.​

A simple query show in the picture helps us understand the data.​

We would prefer 7 day rolling average because Rolling Average are useful for finding long term trends otherwise disguised by occasional fluctuations.

**MySQL Query & Result:**



**Insights:**

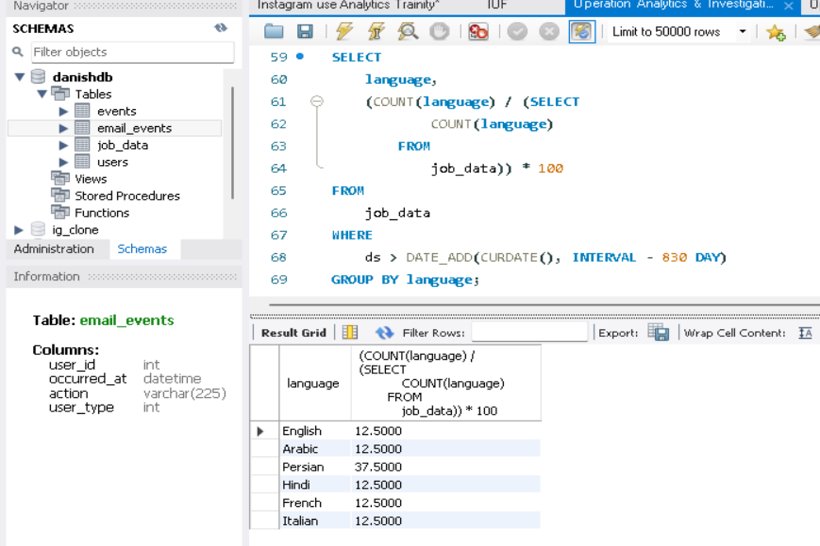
**PERCENTAGE  SHARE OF EACH LANGUAGE:**​

​

We have to find ‘Percentage Share of each Language for last 30 days’.​

The MySQL  query show in the picture gives the data for the percentage share of each language for the last 30 days. ‘Persian’ language has been used the most.

**MySQL Query & Result:**



**Insights:**

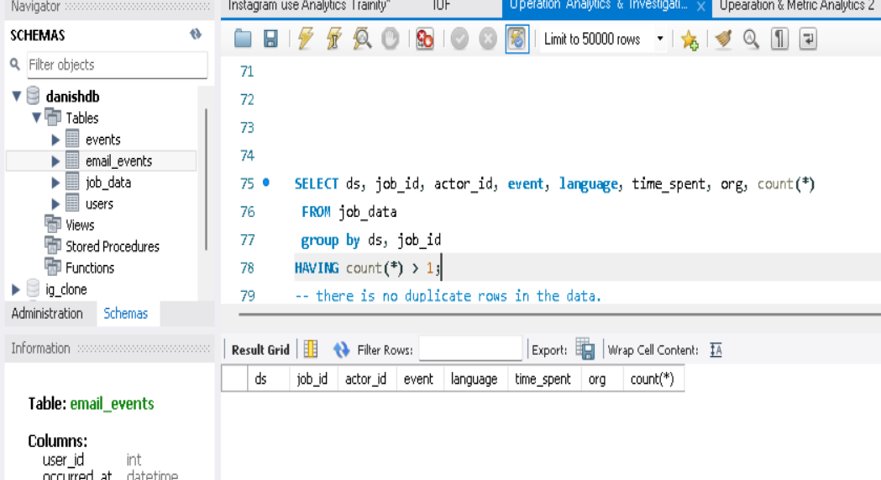
**DUPLICATE ROWS:**​

​

We have to find ‘How to display duplicate rows if there is any’.​

The given table does not have any duplicate rows so no data is displayed but if there is any duplicate data this query from the picture will help in finding it. Duplicate rows is very harmful while deriving insights for future use.

**MySQL Query & Result:**



**Insights:**

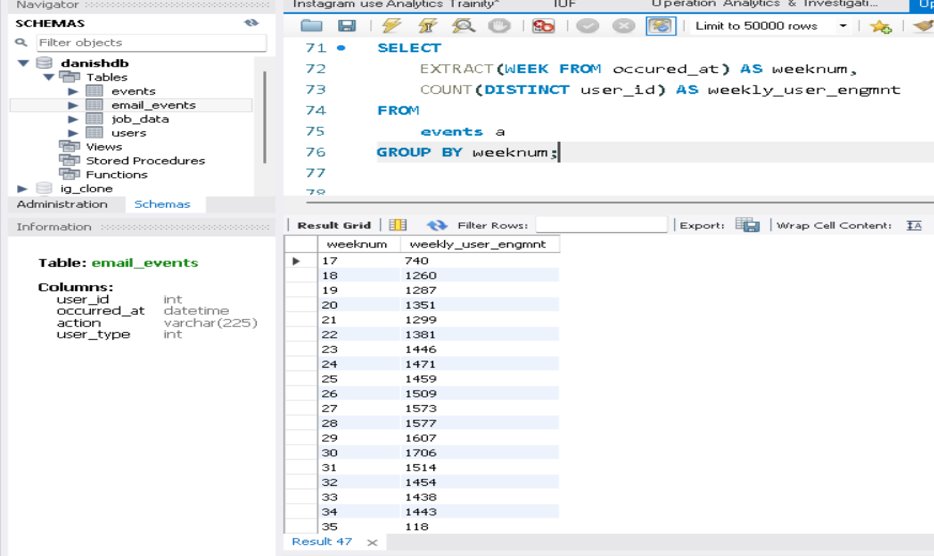
**USER ENGAGEMENT:**​

​

We have to find ‘Weekly User Engagement’. The data will tell us about how many users are engaging with the platform on weekly basis.​

A simple query shown in the picture displays the required data..

**MySQL Query & Result**



**Insights:**

**USER GROWTH:**​

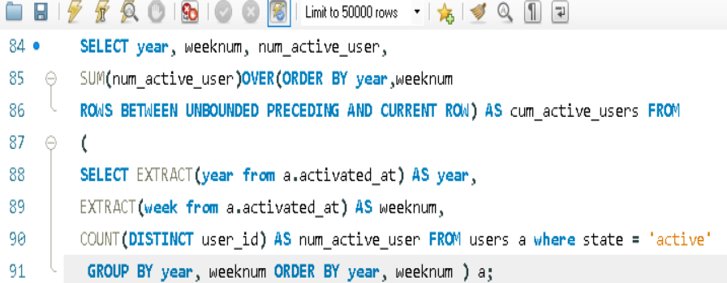
​

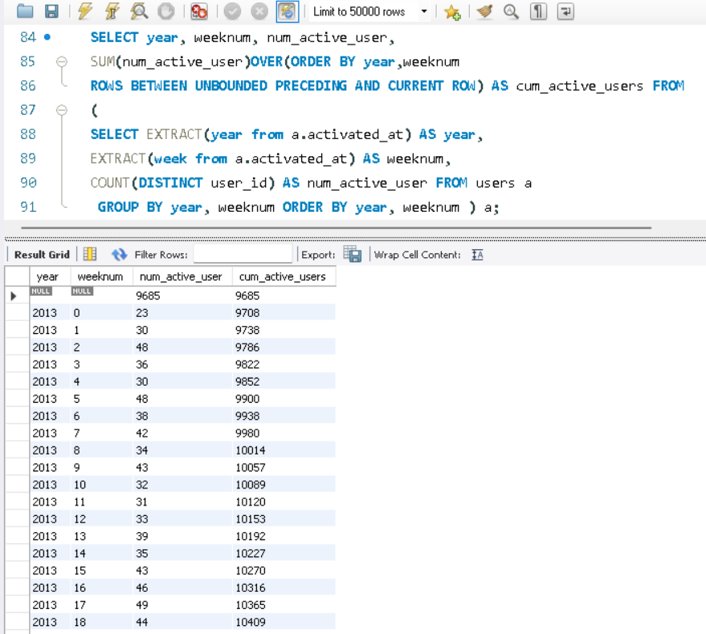
We have to find the ‘User Growth for the Product’ to understand the amount of users growing over time.. ​

​

A simple query has displayed the result as shown in the picture..

**MySQL Query & Result :**





**Insights:**

**WEEKLY RETENTION:**​

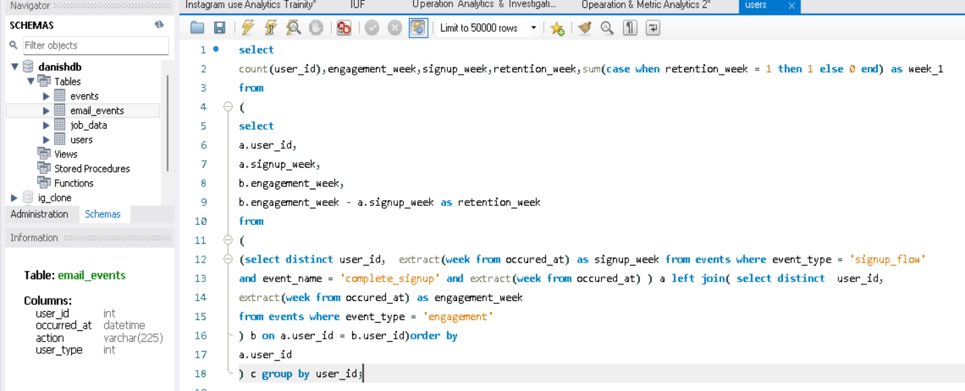
​

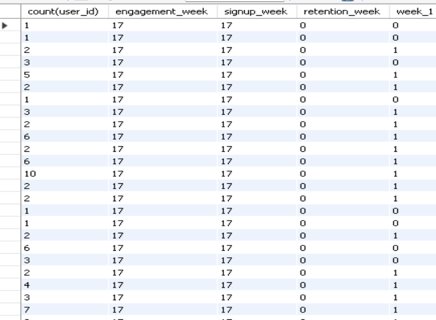
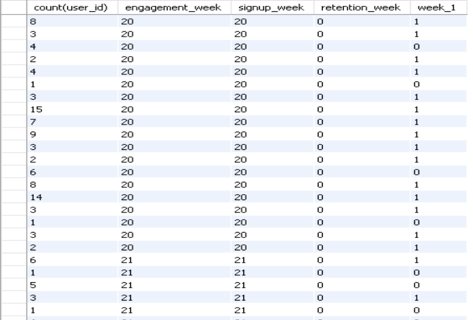
We have to find ‘Weekly Retention of user sign up cohort’ to get the data for users getting retained weekly after signing up for a product . ​

​

To derive a query for this we need to understand ‘COHORT ANALYSIS’ first. The query for this data has been shown in the picture..

**MySQL Query & Result :**



**Insights:**

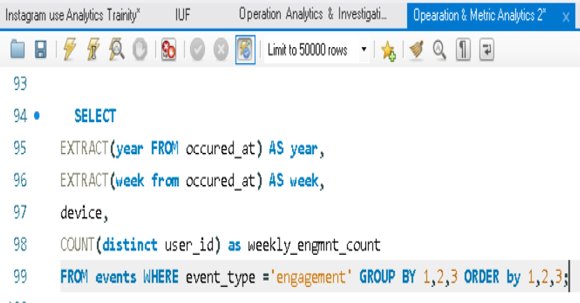
**WEEKLY ENGAGEMENT:**​

​

We have to measure ‘if the user finds quality in the product/services weekly’ to calculate the weekly engagement per device.​

The query and result has been displayed in the picture. The list is very long a part of answer is shown in the picture.

**MySQL Query & Result :**

**Insights:**

**EMAIL ENGAGEMENT:**​

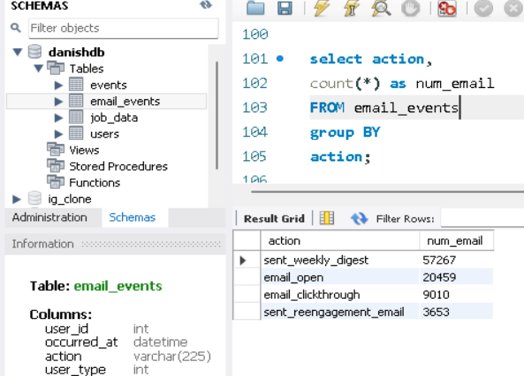
​

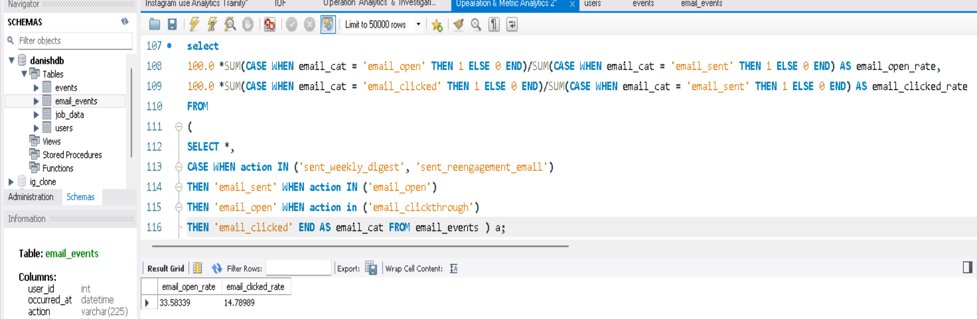
We have to find ‘Email Engagement Metrics’ to get the data for users engaging with email services. ​

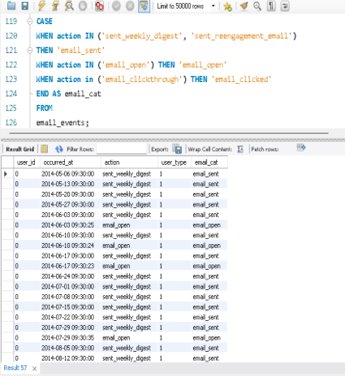
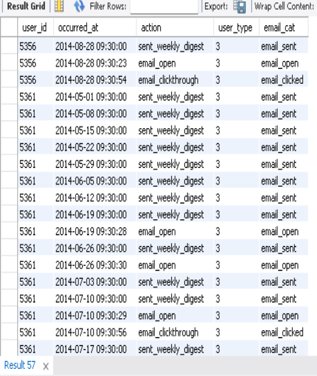
​

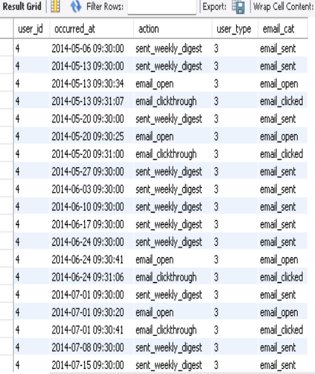
The query and result has been displayed in the picture. List is very long but some answer has only been shown in the picture.

**MySQL Query & Result :**







**Takeaway:**

1. How to get insights about the requirement to run a business.​
2. How to calculate Retention Ratio of users and their engagement on a weekly/yearly/daily basis.​
3. How to run a business using analytics.​s
4. What is Cohort Analysis and how does it work.

**4.Hiring process analytics**

**Project Description:**

Hiring process is the fundamental and the most important function of a company. Here, the MNCs get to know about the major underlying trends about the hiring process. Trends such as- number of rejections, number of interviews, types of jobs, vacancies etc. are important for a company to analyze before hiring freshers or any other individual.

**Approach:**

We are going to Clean the data, by Exploratory Data Analysis and then create charts and graphs to meet the business requirement.

Tech –Stack used

Excel(MS Office Home & Student 2019

**Requirement:**

We are going to find some Insights such as “HIRING”, “AVERAGE SALARY”, “CLASS INTERVAL”, and create different kinds of “CHART/PLOTS” to find some major insights that can help the company grow.

**EXPLORATORY DATA ANALYSIS**

* 1. **Understanding Data Columns & Data:**

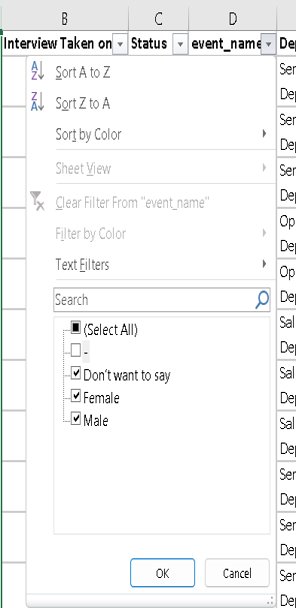
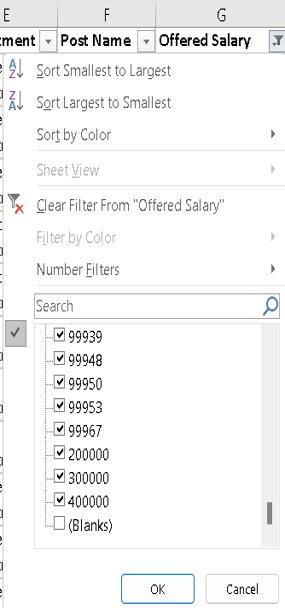
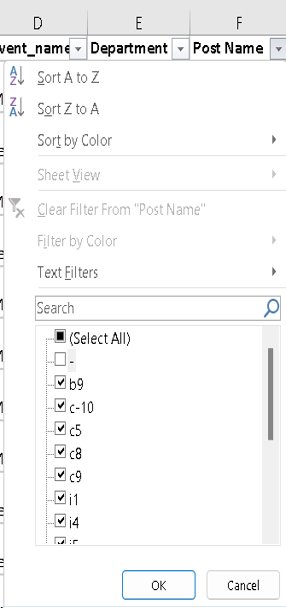
|  |  |
| --- | --- |
| Application\_id​ | Unique Id of each person(Int Datatype). Represented by Numbers​ |
| Interview Taken on​ | The date and time description(Datetime datatype). Datatype format is needed.​ |
| Status​ | Result of Interview(String Datatype). Represented by text​ |
| Event\_name​ | Classified according to gender(Text)​ |
| Department​ | Department for which the interview was taken(Text)​ |
| Post Name​ | Define the post tier(General format). Includes text & numbers denoting a specific code.​ |
| Offered Salary​ | Salary offered for the position for which interview was taken.(Int Datatype). Represented by numbers.​ |

**Checking for Missing Data**

|  |  |
| --- | --- |
| **Column Name​** | **Missing Data if any​** |
| Application\_id​ | No missing data in this column.​ |
| Interview Taken on​ | No missing data.​ |
| Status​ | No missing data.​ |
| Event\_name​ | This column has 15 rows where the data is missing(No values present). ​ |
| Department​ | No missing data.​ |
| Post Name​ | 1 missing data.​ |
| Offered Salary​ | 1 blank row in this column.​ |

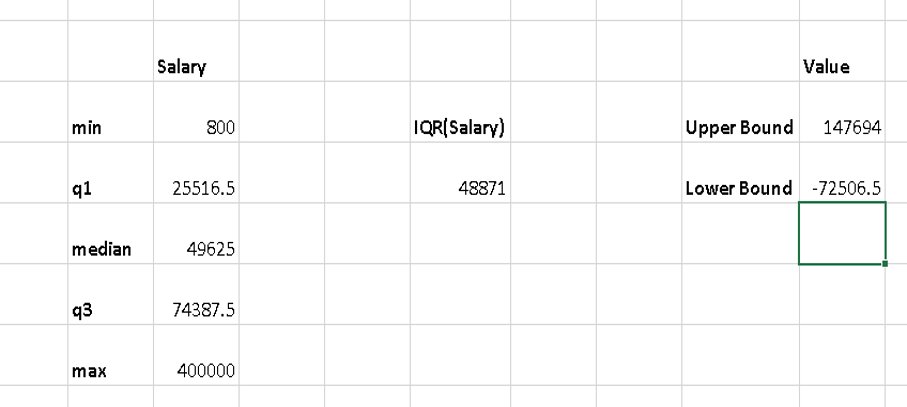
**Data Cleaning & Missing Value Treatment**

1. We have removed those missing rows from the sheet in Excel to make the data more efficient for visualization and Insight Extraction. This process is very important when analyzing any data for a business because Missing Values & Outliers can skew our observations. ​
2. If Missing values are very large in numbers then we cant remove all those rows. In this case we replace these missing values either by MEAN, MEDIAN or any other Data That we Think fits.​
3. In the 6th column where ‘Post Tiers’ data is available the code for each tier is a combination of one letter and a number clubbed together but one of the data category(‘c-10’) contains ‘-’  in between. To make the data look nicer we will remove the ‘-’ to give it a more efficient look.

**Finding Outliers and Removing Them:**

1. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. Outliers can have a disproportionate effect on statistical results, such as the mean, which can result in misleading interpretations.​
2. Her we have used the Quartile and IQR to find if there is an outlier in the Offered Salary column.​
3. We calculated the ‘Upper Bound’ and ‘Lower Bound’ to get the idea about the outlier data.​
4. We found out that there are few outliers in the data and these data has to be removed for better interpretations.​
5. The data has been explained in the picture.



Note: In the given dataset any Offered salary above (147694) & below (-72506.5) is an outlier.  We will remove these rows to get a better and clean dataset.

**Data Description Before and After Cleaning**

|  |  |  |
| --- | --- | --- |
| **​** | **Number of rows Before Cleaning​** | **Number of Rows after Cleaning​** |
| Application\_id​ | 7169​ | 7149​ |
| Interview Taken on​ | 7169​ | 7149​ |
| Status​ | 7169​ | 7149​ |
| Event\_name​ | 7169​ | 7149​ |
| Department​ | 7169​ | 7149​ |
| Post Name​ | 7169​ | 7149​ |
| Offered Salary​ | 7169​ | 7149​ |

These rows had to be removed because they were affecting the Quartile for the given data. If these data were considered they might have skewed the visualization.

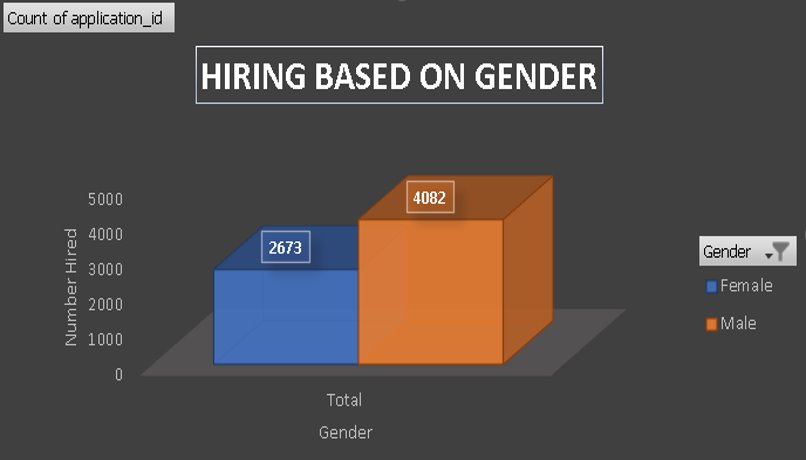
**Visualization & Summary**

**HIRING:**​

**Q.**How Many males &Females are hired?​

​

1. 2673 Females and 4082 Males are hired.​
2. ​
3. It clearly shows that Males are hired almost double that of females**.**



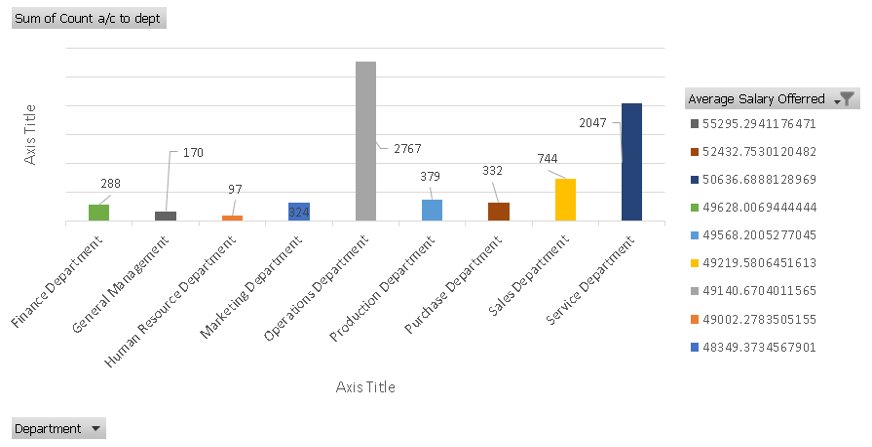
**Visualization & Summary**

**AVERAGE SALARY:**​

**Q.**What is the average salary according to various departments?​

​

1. It is clearly visible that the maximum average salary is for the HR Department followed by Purchase Department.​
2. ​
3. It clearly shows that the least average salary is for the Marketing Department**.**

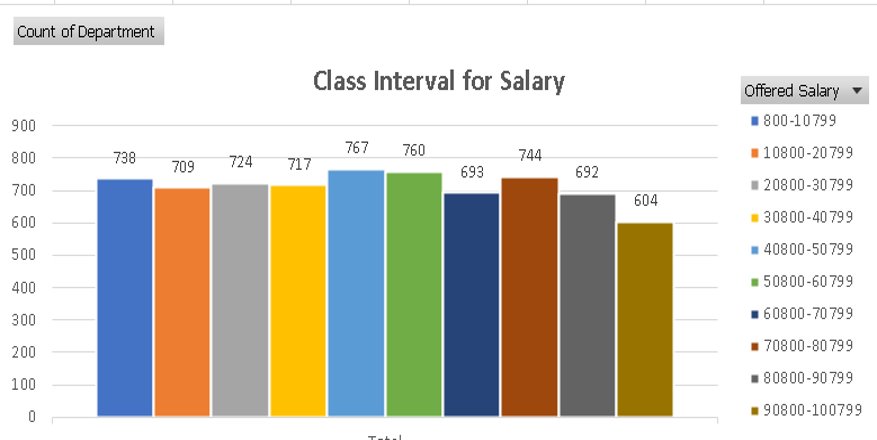


**CLASS INTERVALS:**​

**Q.**Create Class Interval for Salary?​

​

1. A Class Interval of 10000 has been created for Offered Salary to understand how what count of Department falls under which class.​
2. ​
3. The maximum count of department falls under the bracket of 60800-70799

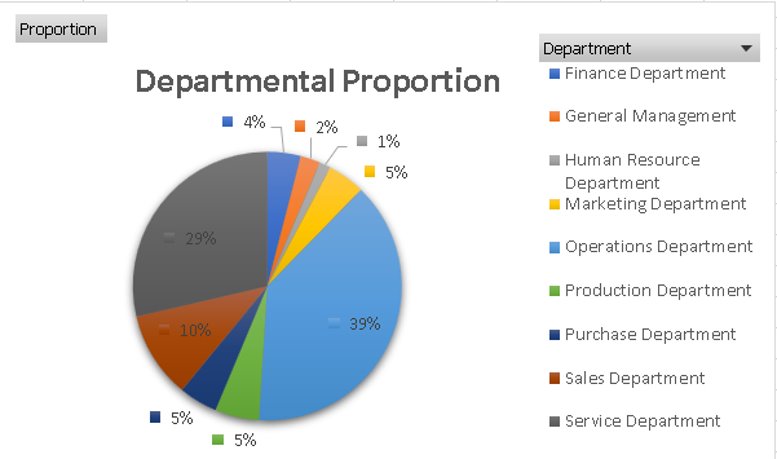


**CHART & PLOTS:**​

**Q.**Draw Pie Chart / Bar Graph ( or any other graph ) to show proportion of people working different department ?​

​

1. Maximum percentage of people work in the Operations Department.​
2. Production & Purchase Department have the same parentage of people working.



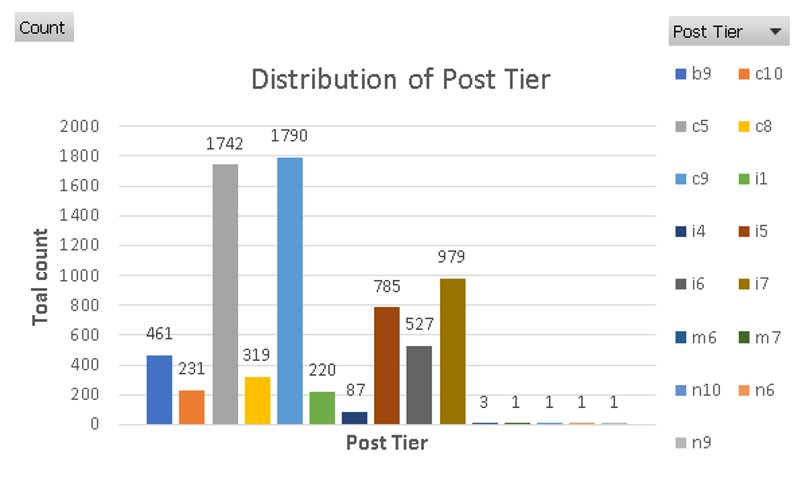
**Visualization & Summary**

**Chart:**​

**Q.**Represent different Post Tier using Chart/Graph?​

​

1. c5 & c9 have the maximum count .​
2. ​
3. M6 to n9 have only 1 candidates  each**.**



**Takeaway:**

1. How to get insights about the requirement to run a business in Excel​
2. How to create class interval.​
3. How to use Quartiles and IQR to find outliers.​
4. Excel visualization

5.IMPACT OF CAR FEATURES

.

5. IMPACT OF CAR FEATURES

**Project description:** As it is familiar that the project is about car features, it is quiet challenging to sort out this .Because the automobile industry has spiked up over the past few decades. The reasons are pretty simple the rapid growth of population, the vehicles have increased rapidly ,fuel usage increased and many new innovations came forward. The demand for cars have also increased among people, due to that the competition also increased among the manufacturers.so, this problem became a headache for the manufacturers.

So, the solution for the problem is pretty simple. We have to use analysis techniques. we have to analyse   pricing, market demand, car features, and should figure out which categories are most familiar for consumers and beneficial for the manufacturers. We have to use regression analysis , market segmentation, so that the producer could develop a efficient pricing strategy that meets consumer demand and profit for the producer as well .for the future development point of view it should be focussed.by doing this, the manufacturer can sustain in the market amidst competition in between.

**Project problem:** In this project,we have to investigate the relationship between car features and its popularity.It will be known by analyzing the popularity variable.By doing so,we can know which car bran d and features are popular among consumers.So that it will be helpful for the manufacturer to make decisions accordingly.

On a complete one,the dataset which is provided has a very huge data and the data is very messy.I have analysed the data and cleaned the data . Cleaning includes removing the duplicate values and removing the null values.

**Approach:** For Data analytics,I have used Microsoft Excel in that I have performed pivot tables, charts,Graphs,Regression etc

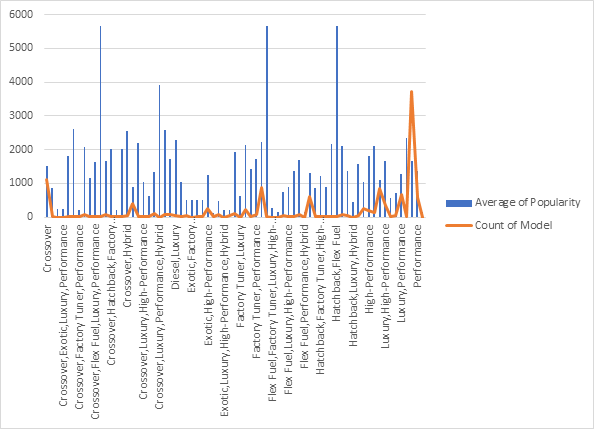
**Tech -Stack used**:Ms-Excel for Data analysis,Ms-Word for Presentation of file.

Tasks:

**Insights required:** How does the popularity of car model vary across different market categories?

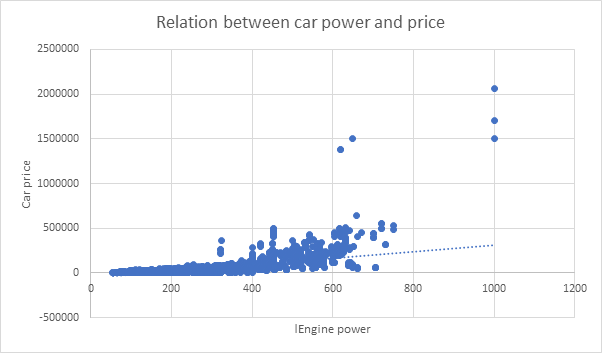
**Task1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

**Task 1.B**:Create a combo chart that visualizes the relationship between  the category and popularity.



**Insights Required:** What is the relationship between a car’s engine power and it’s price?

**Task 2:** Create a scatter chart that plots engine power on x-axis and price will rely on y-axis.Also add a trend line to the scatter chart to visualize the relation between these two variables.

****

**Result**: These two variables have a positive relationship between them. If the Engine Hp increase automatically the price will also increase.

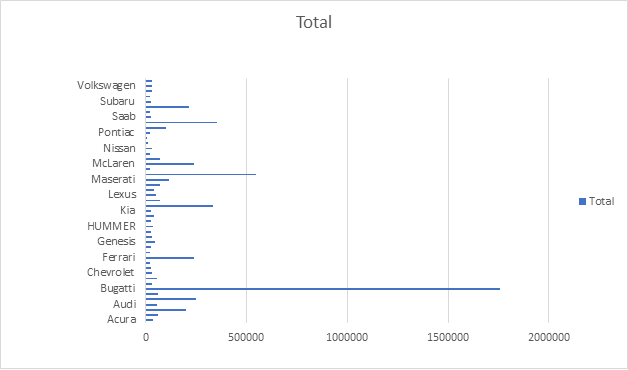
**Insights Required :**Which  car features are most important in determining a car’s price?

**Task 3:** Use the regression analysis to identify the variables which have strongest relationship with the price of a car. Here, a bar chart should be created which shows the coefficient values for each variable to know the importance of them .

**Insight Required:** How does the average price of car vary across different manufacturers.

**Tak 4.A :** Here, create a pivot table that shows average price of cars for each manufacturer.

**Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between the manufacturer and average price.

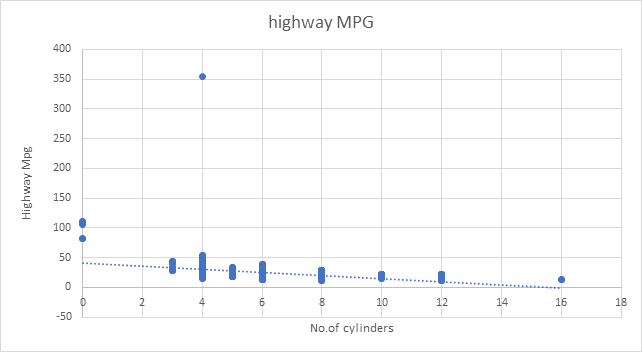
****

**Result**: Bugatti tops the list . It has the highest average price among other car brands.

**Insights required:** What is the relationship between fuel efficiency and number of cylinders in car’s engine?

**Task 5.A:** Create a scatter plot with number of cylinders on the x-axis and highway MPG on the y-axis .Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.

**Task 5.B:** Calculate the correlation  coefficient between the number of cylinders and highway MPG to quantify the strength  and direction of the relationship.



**Result:** Here we can observe that there is a negative relation between No. of cylinders and highway MPG  .If No. of cylinders increases ,highway MPG decreases. We can observe the trendline dropping down.

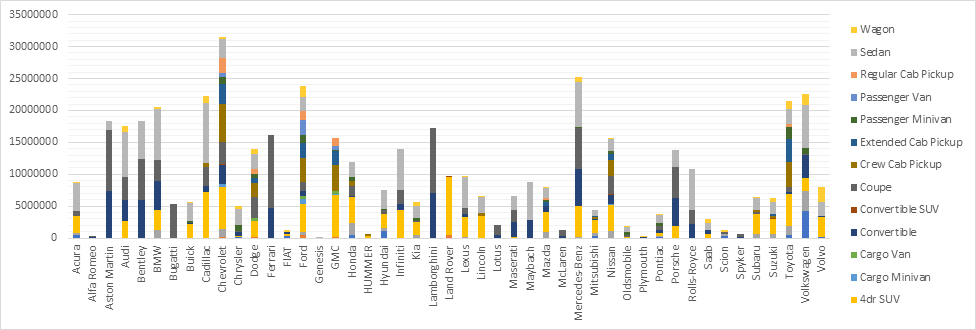
**Dashboard building:**

The further process of the project is creating the dashboard.

Following are the tasks given:

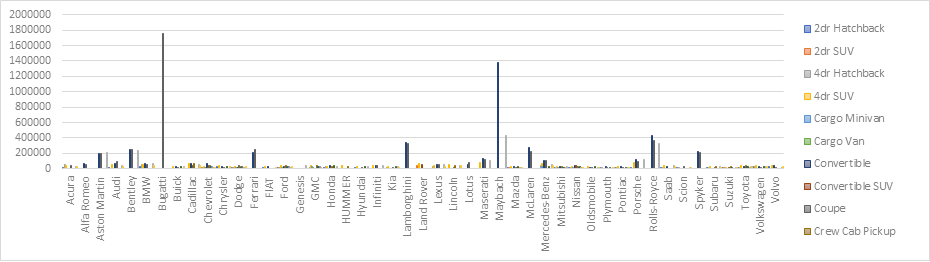
**Task 1:**How does the distribution of car prices vary by brand and body style ?

**Recommended:** Here creating the stacked column chart is recommended.

****

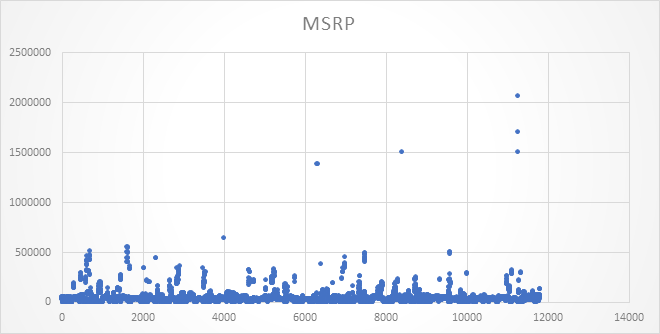
**Result:** By observing this chart,we can conclude that Chevrolet has the highest price distribution.

**Task2:** Which car brands have the highest and lowest average MSRPs,and how does this vary by bodystyle.

****

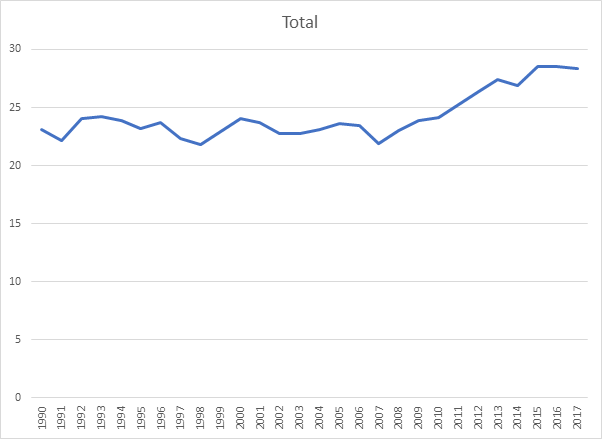
**Result:** Bugatti has the highest MSRP and Plymouth has the lowest average MSRP.

**Task3:**How do the different features such as transmission type affect the MSRP, and how does this vary by body style?

****

**Result:** By seeing this chart,we can conclude that Automatic manual is the most expensive category and the most popular one.

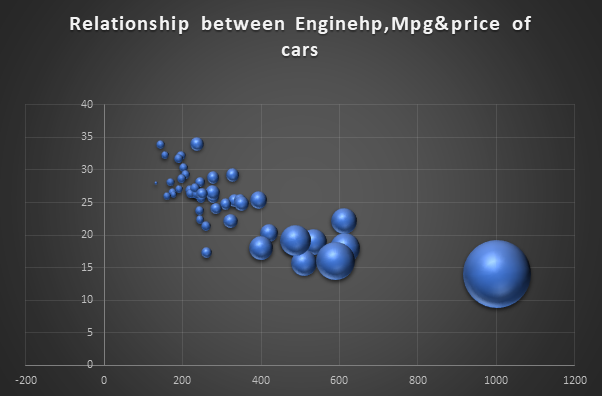
**Task4:**How does the fuel efficiency of cars vary different body styles and model years?



**Result:** As we can observe that there is a slight spike in fuel efficiency over the year.

**Task5:** How do the car’s hp, MPG and price vary across different Brands?

**Recommendation**: Bubble chart Is recommended for this data .



**Result:** We an observe that if engine hp increases highway mpg will decrease and price will also increase.

**Appendix:**

**Appendix:**

**1.Data analytics process:** [**https://1drv.ms/p/s!Am4LmsWQPJYig0ZTg1Jbbampwy5s?e=u9GGz7**](https://1drv.ms/p/s!Am4LmsWQPJYig0ZTg1Jbbampwy5s?e=u9GGz7)

**2. Instagram user Analytics:** [**https://1drv.ms/b/s!Am4LmsWQPJYig2rpozpmbrwSyElU?e=25DHDl**](https://1drv.ms/b/s!Am4LmsWQPJYig2rpozpmbrwSyElU?e=25DHDl)

**3.Operation and metric analysis:** [**https://1drv.ms/p/s!Am4LmsWQPJYihAFyhxHt8kw6FrRo?e=l61I8M**](https://1drv.ms/p/s!Am4LmsWQPJYihAFyhxHt8kw6FrRo?e=l61I8M)

**4. Hiring process analytics:** [**https://1drv.ms/p/s!Am4LmsWQPJYihAQUZSY7EHGK6L3I?e=SYeKhl**](https://1drv.ms/p/s!Am4LmsWQPJYihAQUZSY7EHGK6L3I?e=SYeKhl)

**5. Impact of car features:** [**https://1drv.ms/w/s!Am4LmsWQPJYihFlYC41gUcuVssJR?e=iKoz8p**](https://1drv.ms/w/s!Am4LmsWQPJYihFlYC41gUcuVssJR?e=iKoz8p)