

# **Business Analytics Interview Questions & Answers**



Satyajit Pattnaik

# Agenda

- Business Analyst vs Data Analyst
- Types of questions expected
- Behavioral Questions
- Functional Questions
- Analytical/Brain Teasers
- Technical Questions
  - Theoretical
  - SQL Basics
  - SQL Advanced
  - Python Basics
  - Python Assignment: Hands on
  - Statistical Questions
  - Data Visualization Questions
  - Power BI Questions
  - Basic ML Concepts



# Business vs Data Analysts



**Business Analyst**

Understands and Solves  
a business Problem

Validates business  
requirements

Mediates between the IT  
Team and the  
stakeholders



**Data Analyst**

Gathers and processes  
large data set

Analyzes the collected  
data

Unravels useful business  
insights to the company



**Behavioral  
questions**

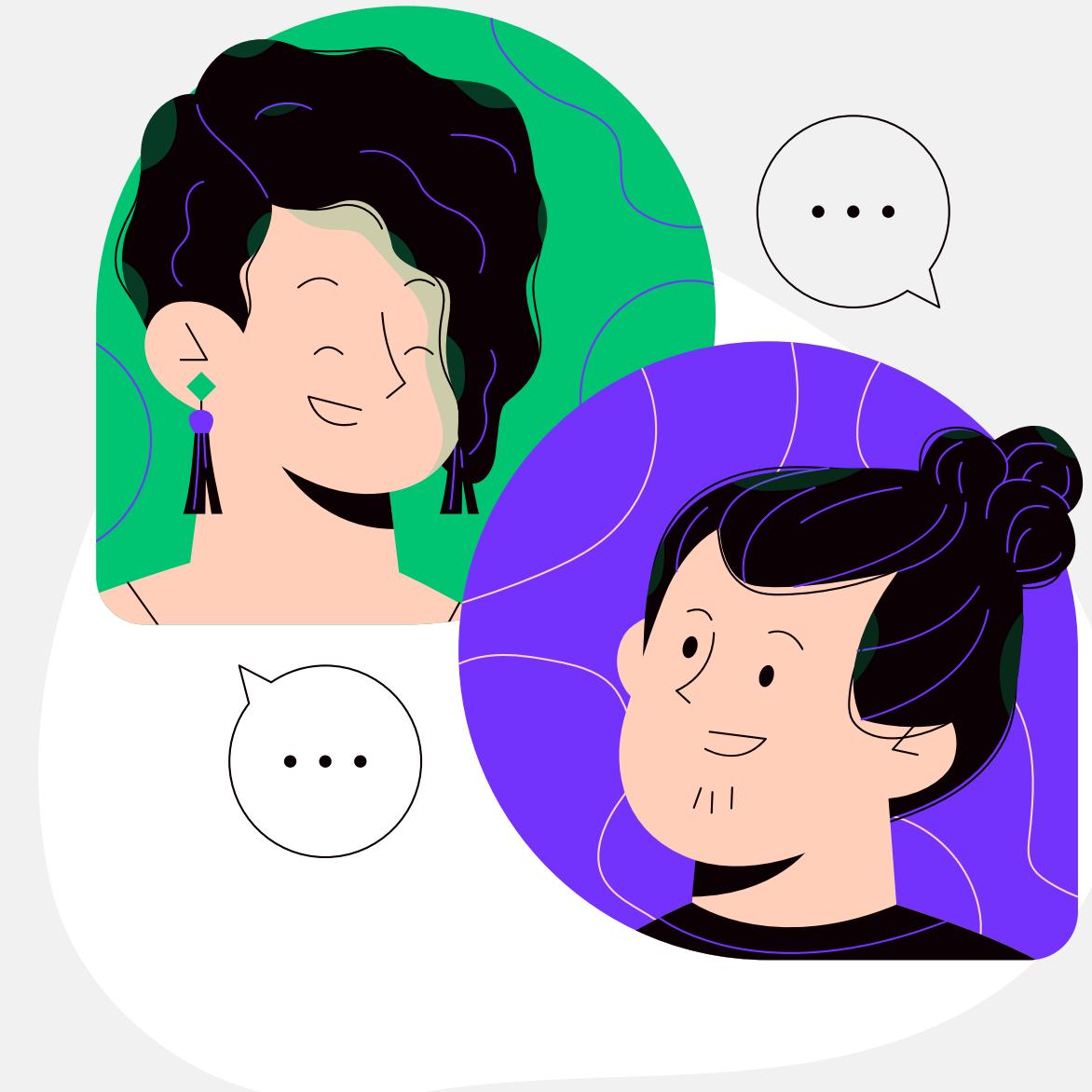
**Technical  
questions**

**Functional  
questions**

**Analytical &  
Scenario  
based  
questions**

# Behavioral questions.

How to handle difficult  
clients & stakeholders

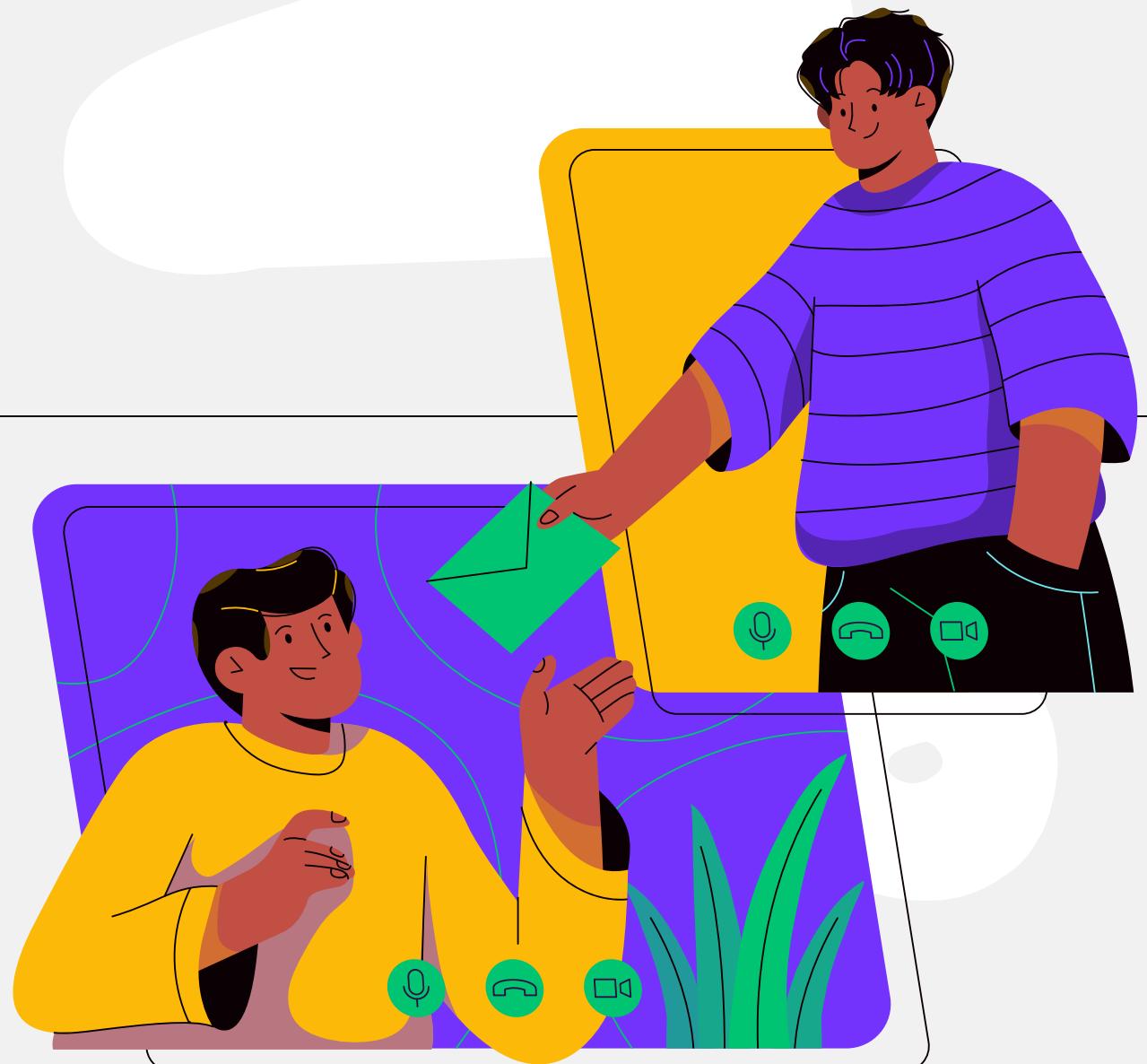


Any mistakes you did  
in past, and how did  
you rectify it

How to pitch an idea  
to the higher  
management, or  
seniors

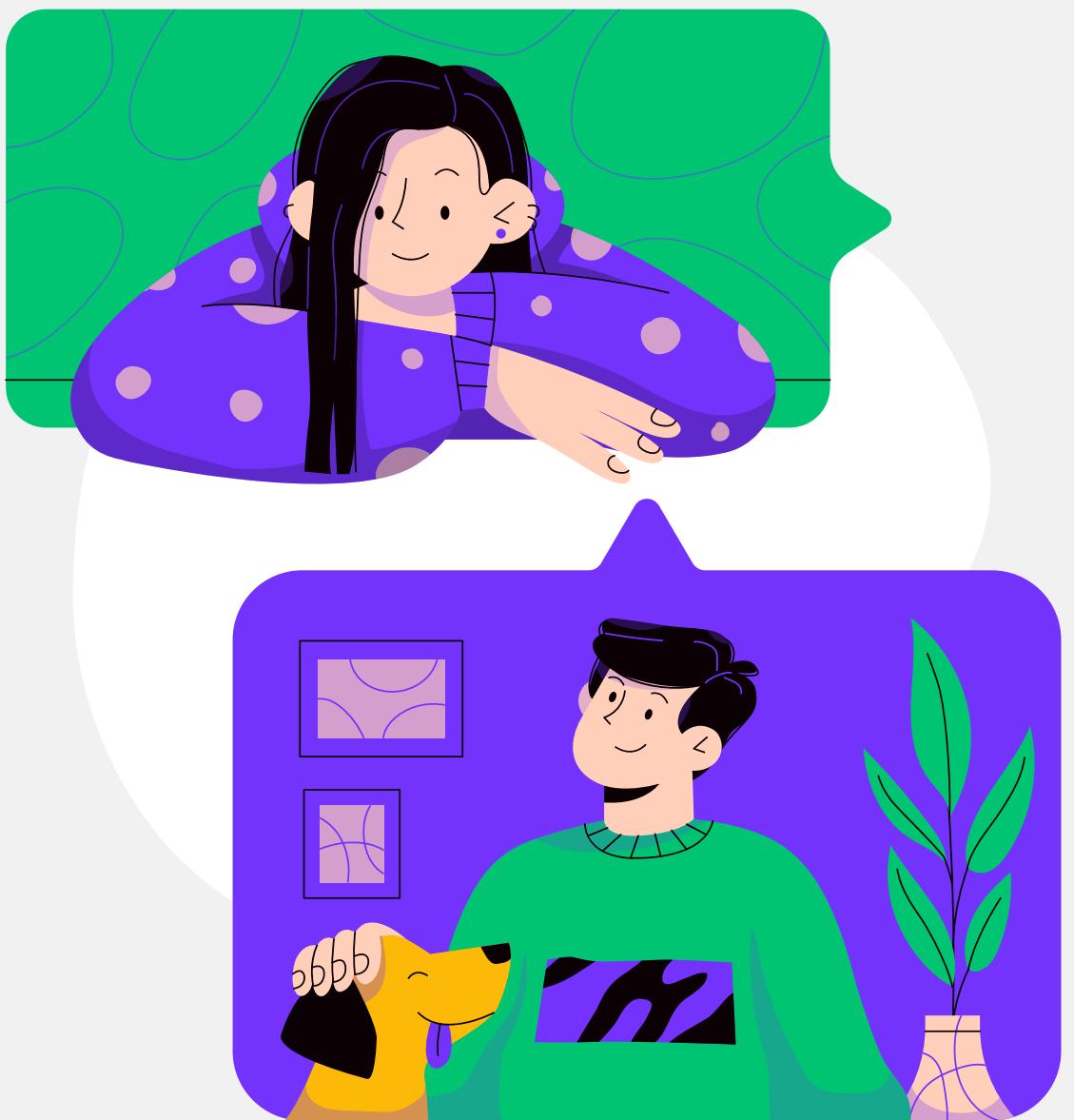
# Functional questions.

- why should we hire you?
- what's your educational background?
- How your typical day looks like?
- where do you see yourself in next 2-5 years?



# Analytical & Scenario based questions.

These are the scenario based questions that are usually asked to play with candidate's brain, and see his analytical power to resolve problems & puzzles



# Puzzle questions.



## Bag of coins

You have 10 bags full of coins, all have infinite coins, and 1 of the 10 bags has fake coins, and you don't remember it. You know that genuine coins weighs 1 gram and fake weighs 1.1 grams. Identify the bag with fake coins in minimum readings.

**Note:** You have a digital weighing machine

# Puzzle questions.



## Shake Hands

At a party, everyone shook hands with everybody else. There were 66 total handshakes, how many were at the party?

# Technical questions.



## Theoretical & Practical

- Basics of Python, & Statistics
- SQL, Advanced SQL
- Dashboard Building
- Power BI/Tableau
- Basics of Machine Learning & AI

# Theoretical Questions

11

Video Reference: [Click here](#)

## Types of SQL Statements

- SQL definition: DDL: Create, alter, & drop
- SQL manipulation: DML: Delete, Insert, Select, Update
- SQL control: Transaction Control Statements: Commit, Rollback, Set Transaction, Session Control: Alter Session, Set Role etc.

## Experience with technical & functional documents

BA are responsible for the documentation, and the initial requirement gathering about the project, so they need to create documents such as Stakeholder Analysis, Scope Statement, and if requirement, they also need to work on Competitor Analysis.

## How do you convey complex, technical information to non-tech stakeholders

The way you answer will showcase your communication skills, so you have to be relatable, and able to create simple mockups and answers, so storytelling skills is going to be tested here.

# Theoretical Questions

Describe your experience with UAT?

- Planning
  - Execution
  - Documentation
  - Evaluation
  - Reporting & Lessons Learned
- 

PaaS, SaaS, IaaS

- Platform as a service: allows developers to build apps over the internet
  - Software as a service: 3rd party to host applications & give access
  - Internet as a service: form of cloud computing that provides virtual computing resources through the internet
- 

After researching, you come across two possible solutions, one is cloud based, and other is on-premises, which one would you recommend and why?

No concrete answer....

# Theoretical Questions

which visualization tools have you used

- Power BI
  - Tableau
  - QlikSense
  - Alteryx
  - MS Excel
- 

How to build a predictive model?

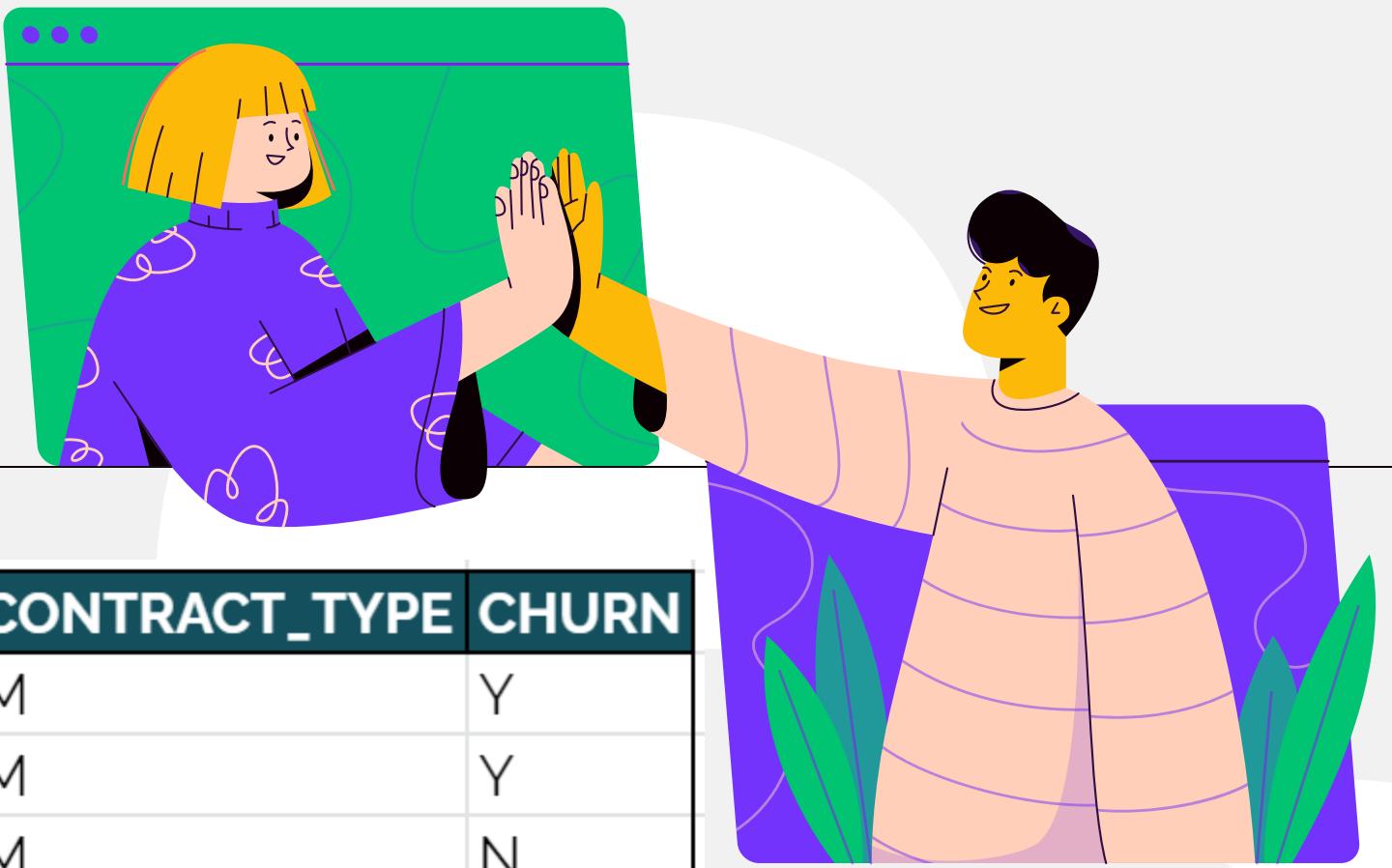
Give examples and explain, and be ready to tackle related questions to the use case you explain.

---

Explain one of the ML use cases you worked on

Based on your resume

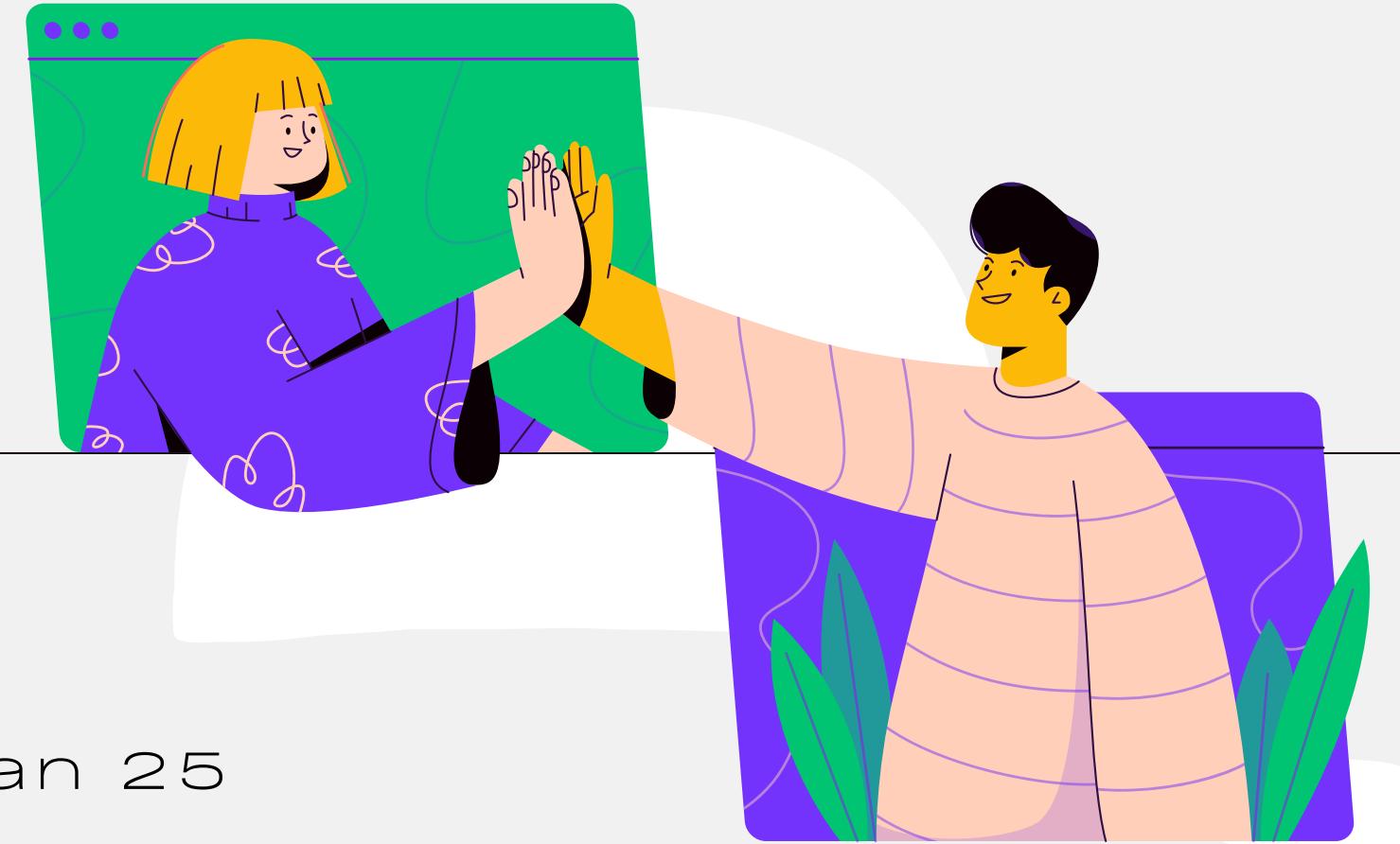
# SQL Basics.



CUSTOMER #	FIRST_NAME	LAST_NAME	GENDER	AGE	CITY	CONTRACT_TYPE	CHURN
1	Satyajit	Pattnaik	M	31	Delhi	M	Y
2	Saurav	Agarwal	M	31	Mumbai	M	Y
3	Sudhanshu	Kumar	M	30	Delhi	M	N
4	Ravi	Kumar	M	35	Hyderabad	Y	N
5	Sukhwinder	Singh	M	29	Delhi	Y	Y
6	Madhumita	Banerjee	F	12	Hyderabad	Y	N
7	Raj	Kumar	M	15	Delhi	Y	N
8	Namrata	Nayak	F	13	Delhi	2Y	N
9	Reeta	Pattnaik	F	29	Mumbai	2Y	N
10	Janvi	Rath	F	18	Mumbai	2Y	N

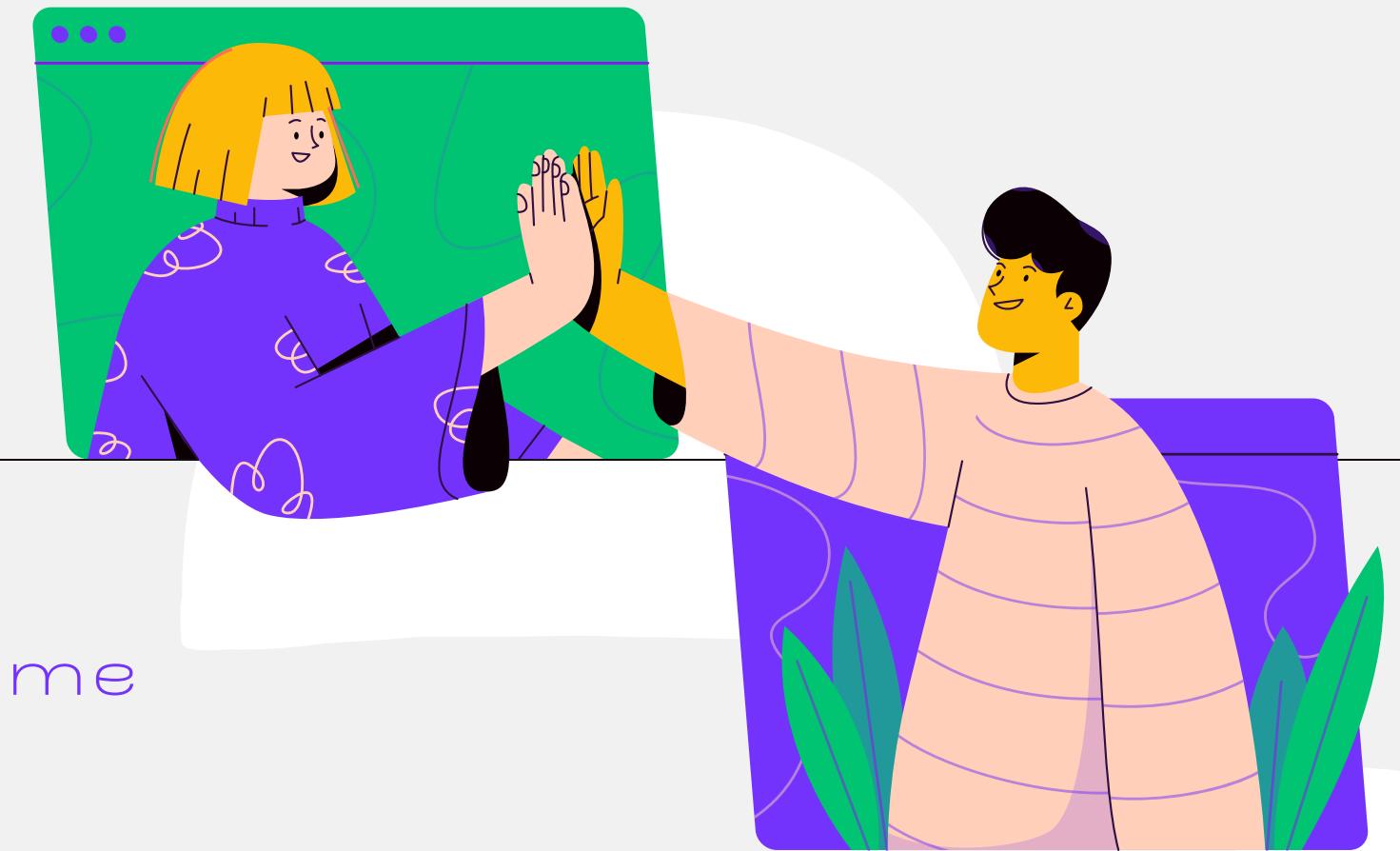
# SQL

## Basics.



- Retrieve all the records from a table
- Retrieve two columns from a table
- Retrieve customers having age more than 25
- Retrieve customers having age more than 25 and from Mumbai
- Retrieve customers having age more than 25 OR churn: Y
- Retrieve customers having age between 15 and 35
- Retrieve customers staying in Delhi & Mumbai
- Find the total number of records in the table
- Insert a customer to the table
- Update the customer details as churned whose customer id is 7
- Find the minimum and maximum age from the list of customers

# Answers



- select \* from tablename
- select column A, column B from tablename
- select \* from tablename where age > 25
- select \* from tablename where age > 25 and city = 'Mumbai'
- select \* from tablename where age > 25 OR churn = 'Y'
- select \* from tablename where age between 15 and 35
- select \* from tablename where city in ('Delhi', 'Mumbai')
- select count(\*) from tablename
- Insert into tablename values ()
- update tablename set churn = 'Y' where cust\_id = 7
- select \* from tablename where age in (select min(age) from tablename)



# SQL

## Advanced.

Facebook



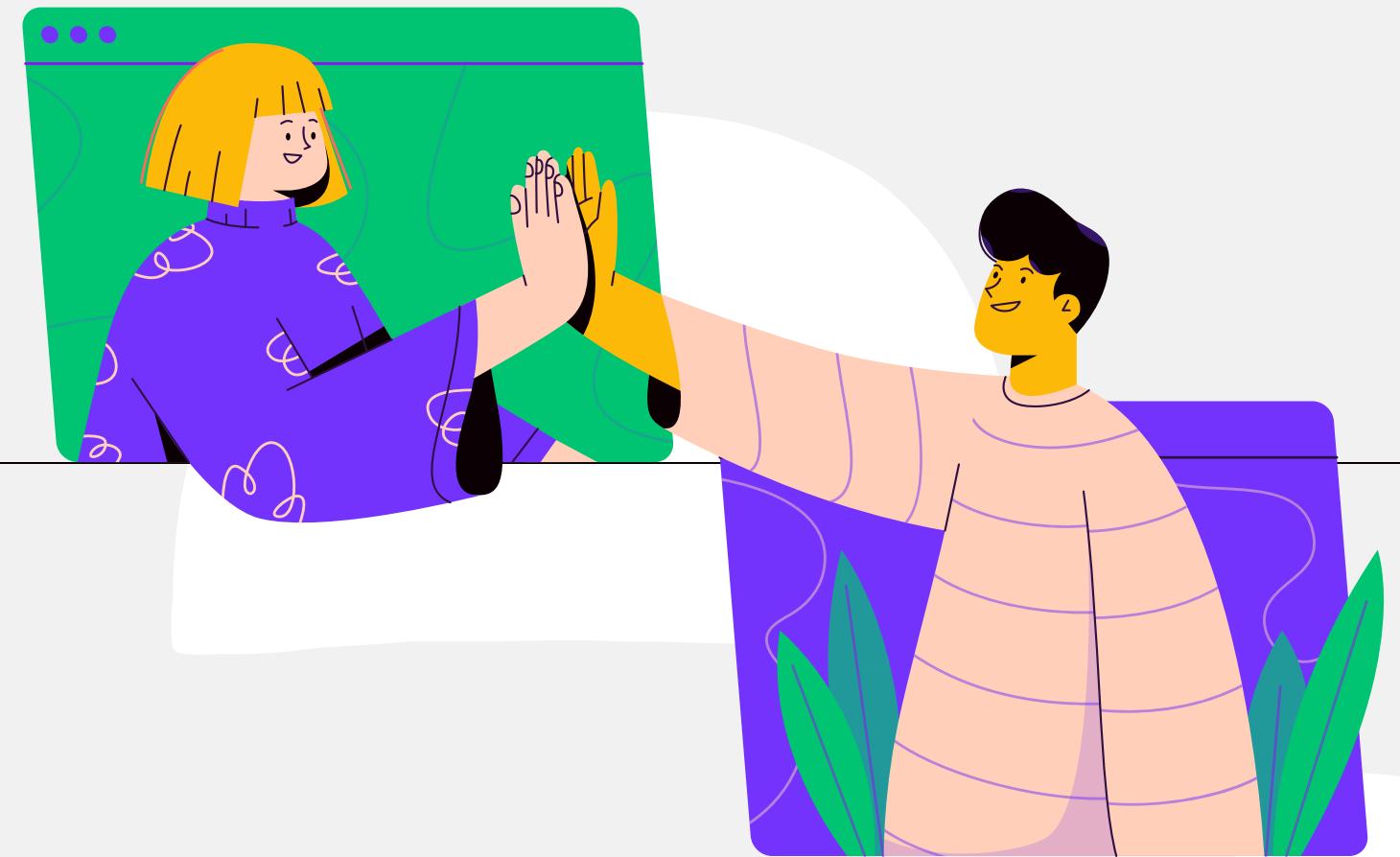
Let say, you have **date**, **wifi id**, **wifi speed**, **latency**, **country**

1. Calculate the average download speed per wifi for a particular date
2. same question, show it for last 7 days

# SQL

# Advanced.

## Answers



1. select avg(wifi\_speed), wifi\_id from tablename where date = 'specified\_date' group by wifi\_id
2. select avg(wifi\_speed), wifi\_id from tablename where date >= date\_add(SYSDATE() interval-7 day) group by wifi\_id order by date desc

# SQL Advanced

## Facebook

write a SQL query to find the cancellation rate of requests with unbanned users (both client and driver must not be banned) each day between "2013-10-01" and "2013-10-03".

The cancellation rate is computed by dividing the number of canceled [by client or driver] requests with unbanned users by the total number of requests with unbanned users on that day.

Return the result table in any order. Round Cancellation Rate to two decimal points.



Table: Trips

Column Name	Type
<code>Id</code>	<code>int</code>
<code>Client_Id</code>	<code>int</code>
<code>Driver_Id</code>	<code>int</code>
<code>city_Id</code>	<code>int</code>
<code>Status</code>	<code>enum</code>
<code>Request_at</code>	<code>date</code>

`Id` is the primary key for this table.

The table holds all taxi trips. Each trip has a unique `Id`, while `Client_Id` and `Driver_Id` are foreign keys to the `Users_Id` at the `Users` table.

`Status` is an ENUM type of ('completed', 'cancelled\_by\_driver', 'cancelled\_by\_client').

Table: Users

Column Name	Type
<code>Users_Id</code>	<code>int</code>
<code>Banned</code>	<code>enum</code>
<code>Role</code>	<code>enum</code>

`Users_Id` is the primary key for this table.

The table holds all users. Each user has a unique `Users_Id`, and `Role` is an ENUM type of ('client', 'driver', 'partner').

`Status` is an ENUM type of ('Yes', 'No').



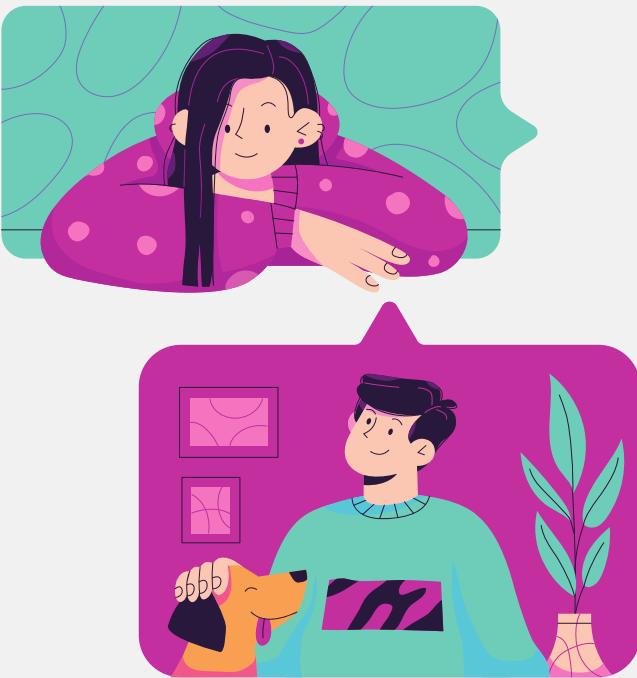
# SQL Advanced

## Facebook

Source:  
Leetcode

### Solution

```
with r1 as [
    select sum(case when t.status like 'cancelled%' then 1 else 0 end)
    as sum_canc, count(*) as tot_trips, t.request_at
    from trips t
    where 1=1
        and client_id not in (select users_id from users where banned =
        'Yes')
        and driver_id not in (select users_id from users where banned =
        'Yes')
        and t.request_at between '2013-10-01' and '2013-10-03'
    group by 3
]
select r1.request_at as Day, round(sum_canc/tot_trips,2) as
"Cancellation Rate" from r1
```



# SQL Advanced

## Facebook

A company's executives are interested in seeing who earns the most money in each of the company's departments. A high earner in a department is an employee who has a salary in the top three unique salaries for that department.

Write an SQL query to find the employees who are high earners in each of the departments.

Return the result table in any order.

Table: Employee

Column Name	Type
<code>Id</code>	<code>int</code>
<code>Name</code>	<code>varchar</code>
<code>Salary</code>	<code>int</code>
<code>DepartmentId</code>	<code>int</code>

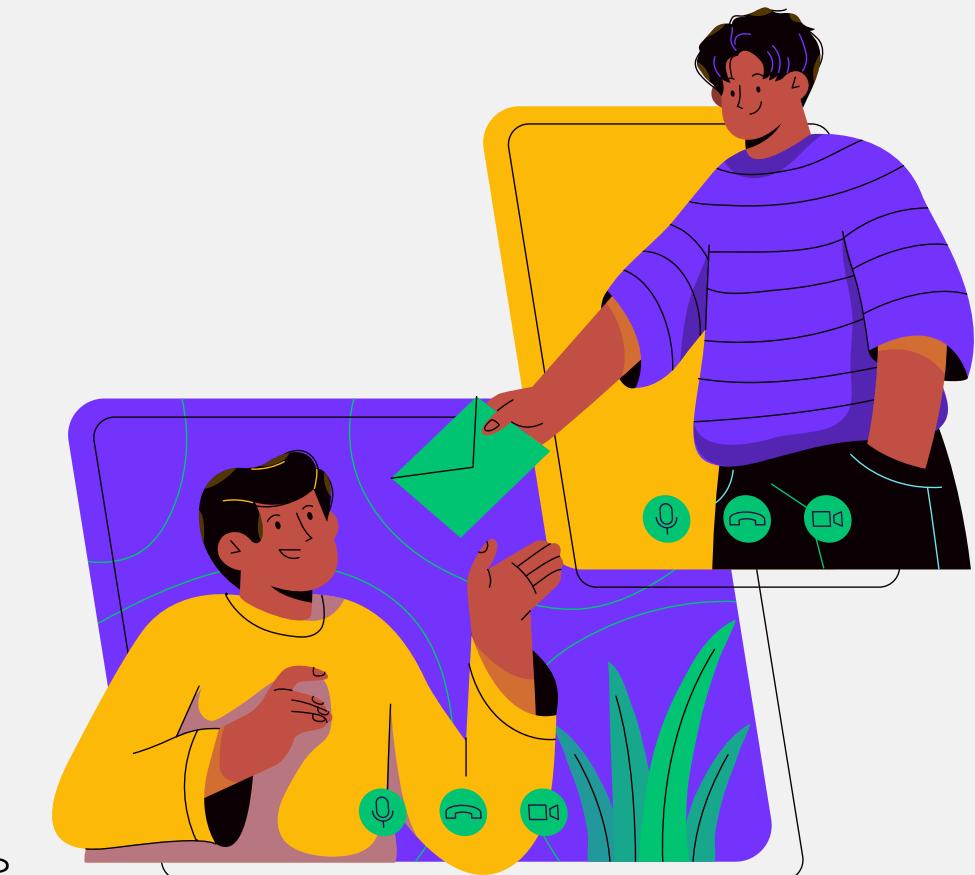
`Id` is the primary key for this table.  
Each row contains the ID, name, salary, and department of one employee.

Table: Department

Column Name	Type
<code>Id</code>	<code>int</code>
<code>Name</code>	<code>varchar</code>

`Id` is the primary key for this table.  
Each row contains the ID and the name of one department.

Source:  
Leetcode

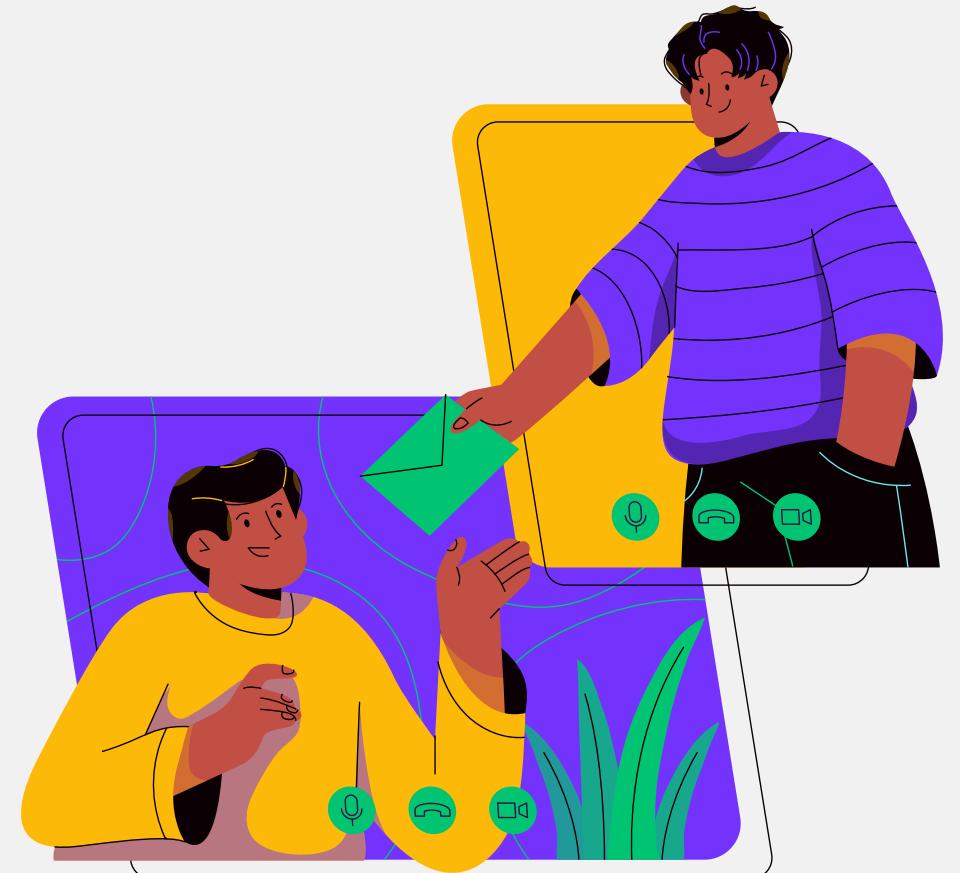


# SQL Advanced

## Facebook

```
SELECT
    d.Name AS 'Department', e1.Name AS 'Employee',
    e1.Salary
FROM
    Employee e1
    JOIN
    Department d ON e1.DepartmentId = d.Id
WHERE
    3 > (SELECT
            COUNT(DISTINCT e2.Salary)
        FROM
            Employee e2
        WHERE
            e2.Salary > e1.Salary
            AND e1.DepartmentId = e2.DepartmentId
    )
;
;
```

Source:  
Leetcode



# SQL Advanced

## Facebook

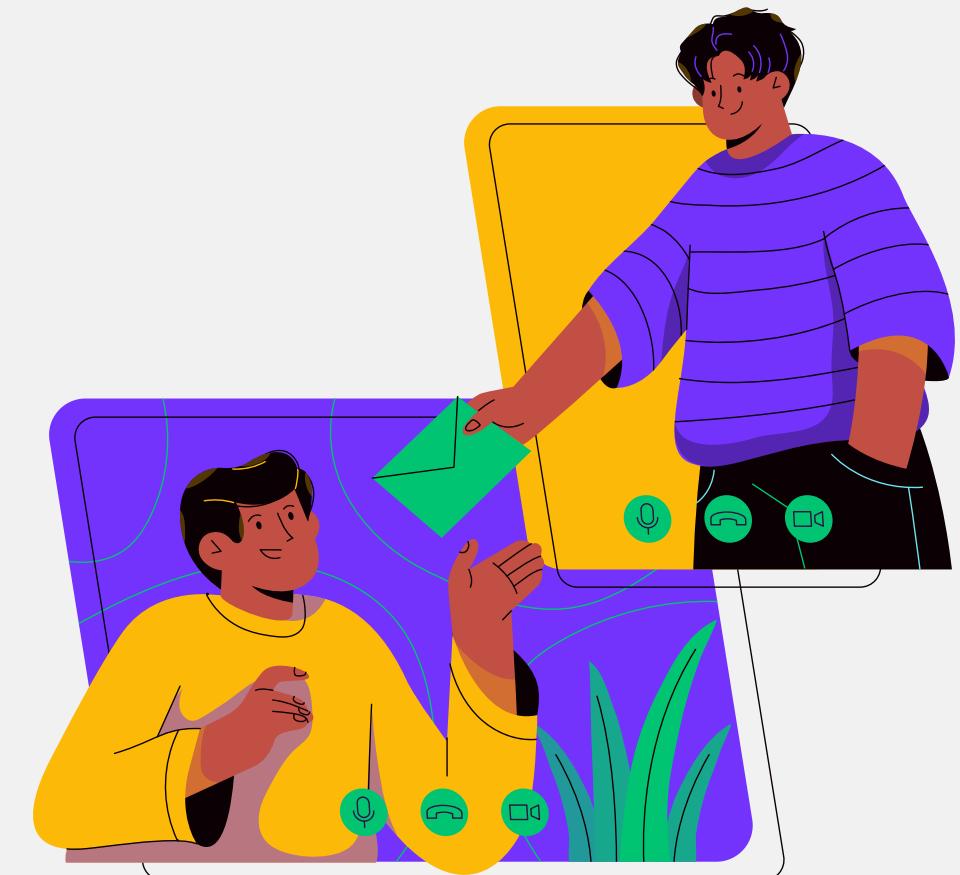
Write a SQL query to get the nth highest salary from the Employee table.

Id	Salary
1	100
2	200
3	300

For example, given the above Employee table, the nth highest salary where n = 2 is 200. If there is no nth highest salary, then the query should return null.

getNthHighestSalary(2)
200

Source:  
Leetcode

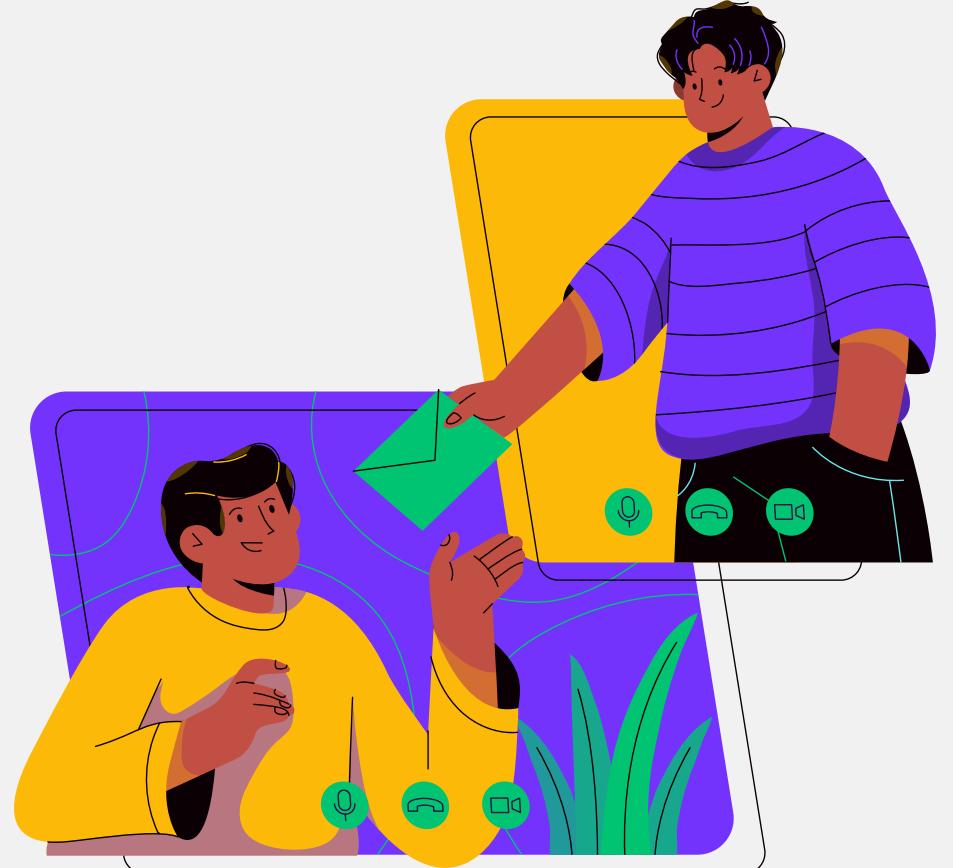


# SQL Advanced

Facebook

Source:  
Leetcode

```
CREATE FUNCTION getNthHighestSalary(N INT)
RETURNS INT
BEGIN
RETURN (
# Write your MySQL query statement below.
select DISTINCT salary as getNthHighestSalary FROM
(select dense_rank() over (order by salary desc) rnk,
salary from employee) x
WHERE X.RNK=N
];
END
```



# SQL Advanced

## Facebook

Source:  
Leetcode

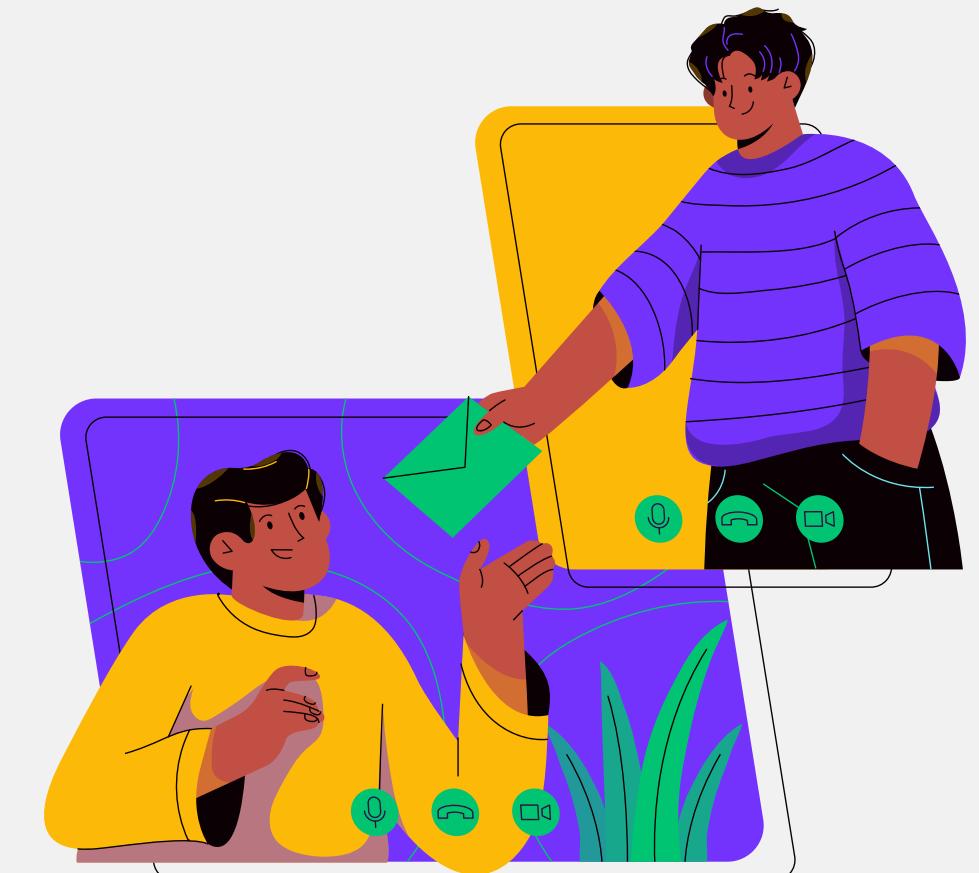
Write a SQL query to find employees who have the highest salary in each of the departments.

The Employee table holds all employees.  
Every employee has an Id, a salary, and  
there is also a column for the department Id.

Id	Name	Salary	DepartmentId
1	Joe	70000	1
2	Jim	90000	1
3	Henry	80000	2
4	Sam	60000	2
5	Max	90000	1

The Department table holds all departments of the company.

Id	Name
1	IT
2	Sales

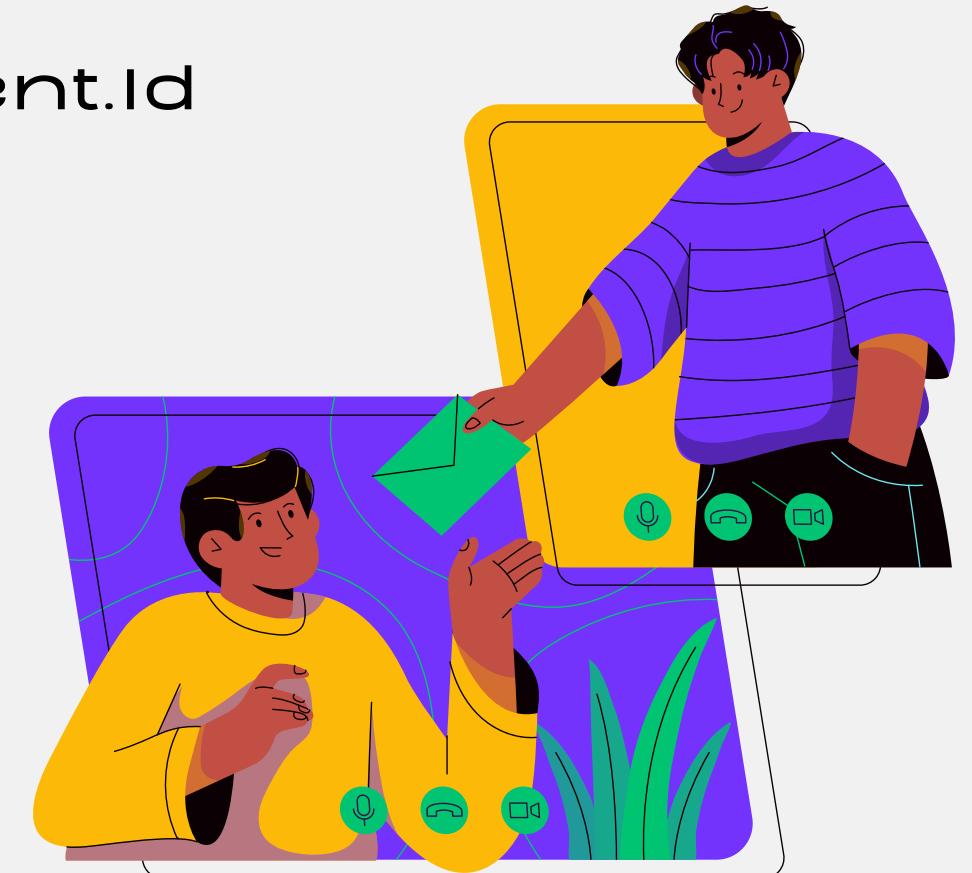


# SQL Advanced

## Facebook

Source:  
Leetcode

```
SELECT
    Department.name AS 'Department',
    Employee.name AS 'Employee',
    Salary
FROM
    Employee
    JOIN
    Department ON Employee.DepartmentId = Department.Id
WHERE
    (Employee.DepartmentId , Salary) IN
    (  SELECT
        DepartmentId, MAX(Salary)
    FROM
        Employee
    GROUP BY DepartmentId
)
;
;
```



# SQL Advanced

## Facebook

Source:  
Leetcode

Write an SQL query to find all numbers that appear at least three times consecutively.

Return the result table in any order

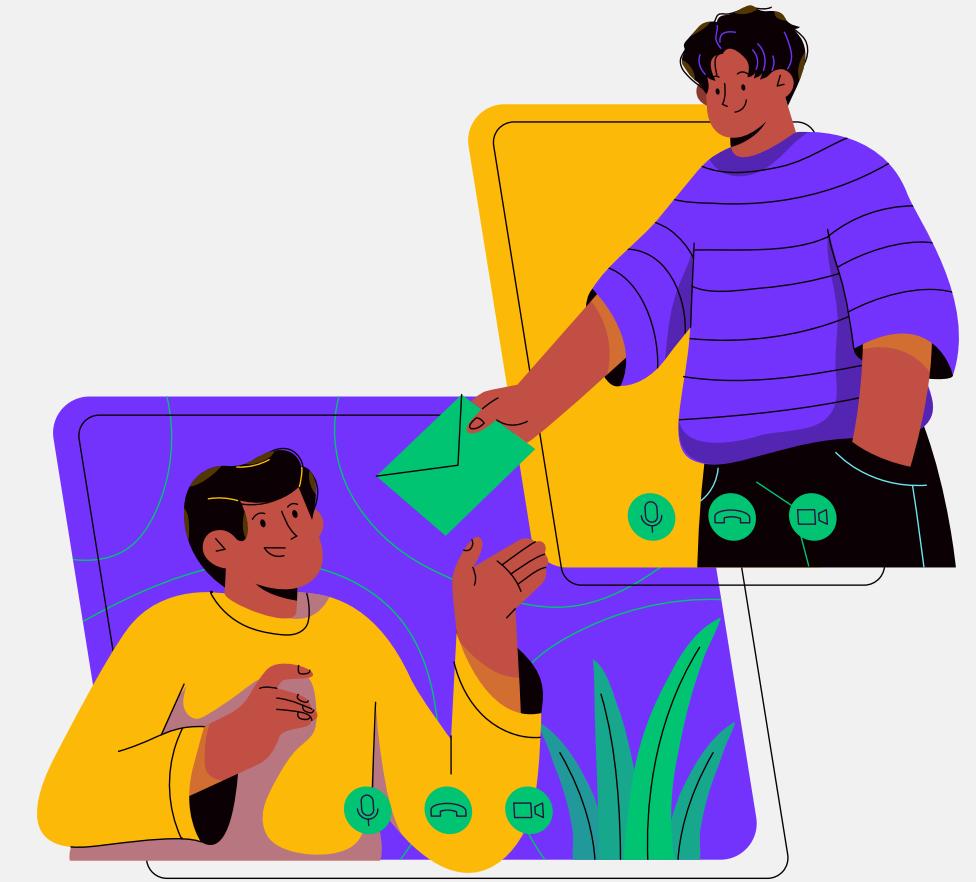
Table: Logs

Column Name	Type
id	int
num	varchar

id is the primary key for this table.

### Solution

```
SELECT DISTINCT
    l1.Num AS ConsecutiveNums
FROM
    Logs l1,
    Logs l2,
    Logs l3
WHERE
    l1.id = l2.id - 1
    AND l2.id = l3.id - 1
    AND l1.Num = l2.Num
    AND l2.Num = l3.Num
;
```



# SQL Advanced

## Facebook

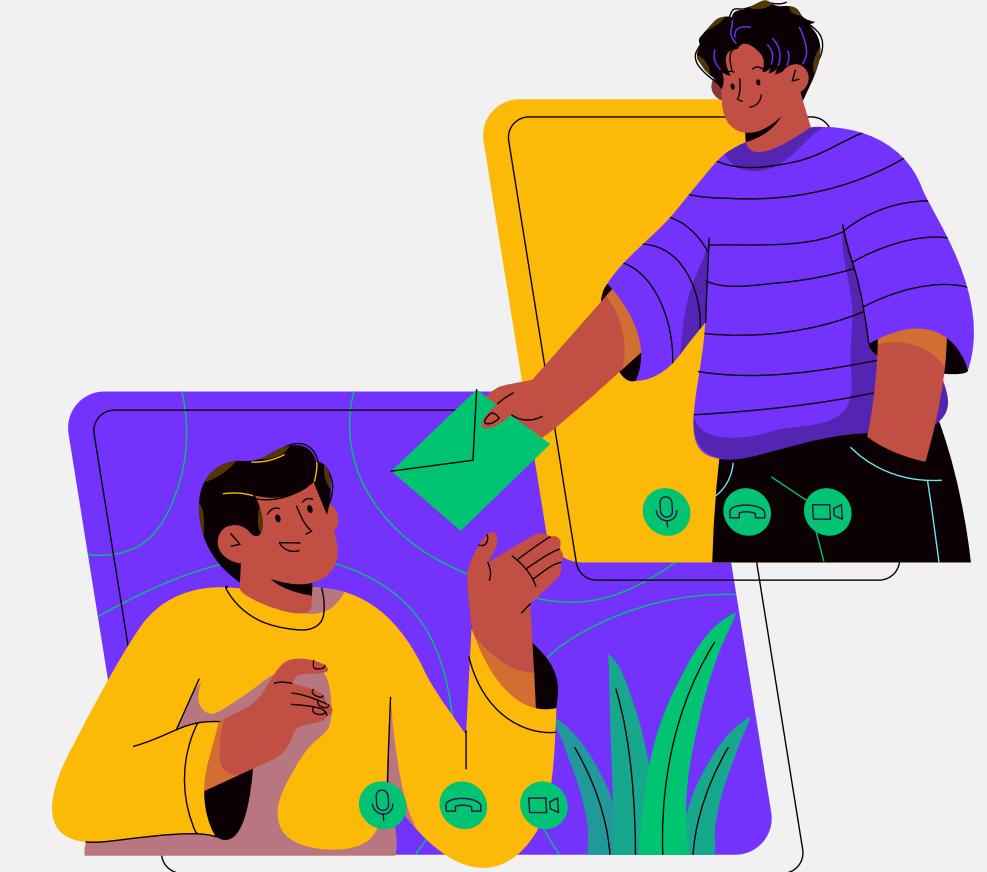
Source:  
Leetcode

Write a SQL query to rank scores. If there is a tie between two scores, both should have the same ranking. Note that after a tie, the next ranking number should be the next consecutive integer value. In other words, there should be no "holes" between ranks.

Id	Score
1	3.50
2	3.65
3	4.00
4	3.85
5	4.00
6	3.65

Solution:

```
select
score,dense_rank()over[order
by score desc] as 'Rank'
from Scores
order by 'Rank'
```



# SQL Advanced

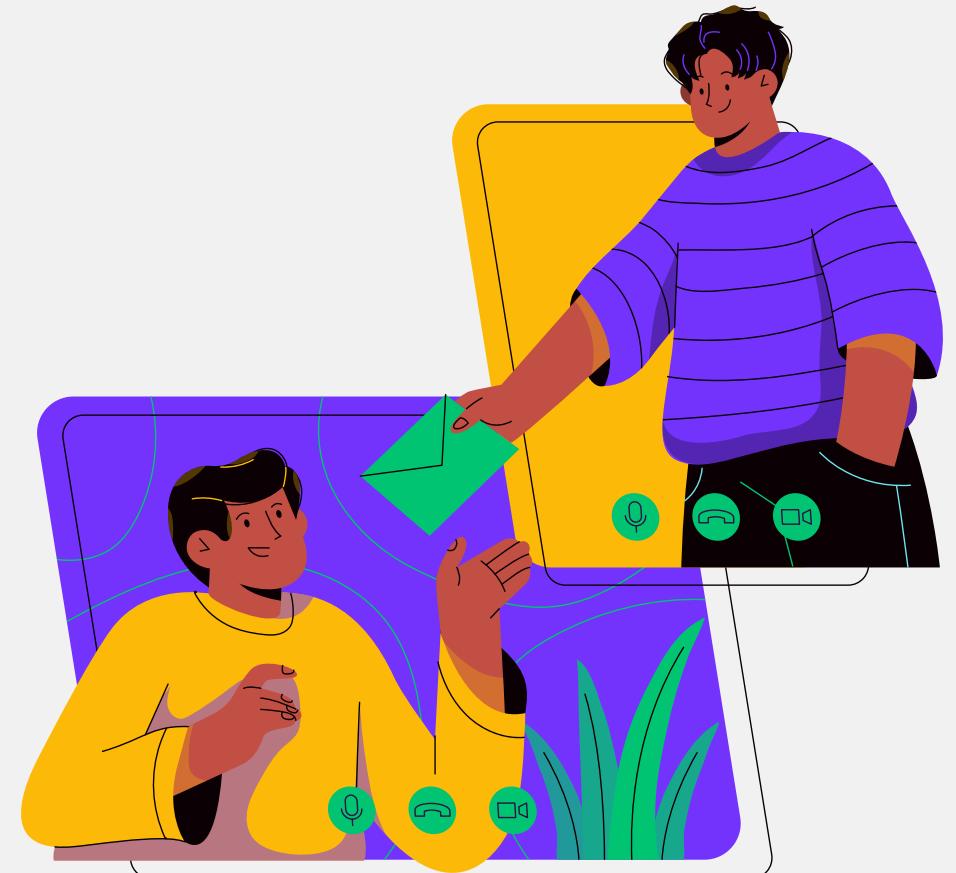
## Facebook

Write a SQL query to get the second highest salary from the Employee table.

Id	Salary
1	100
2	200
3	300

### Solution

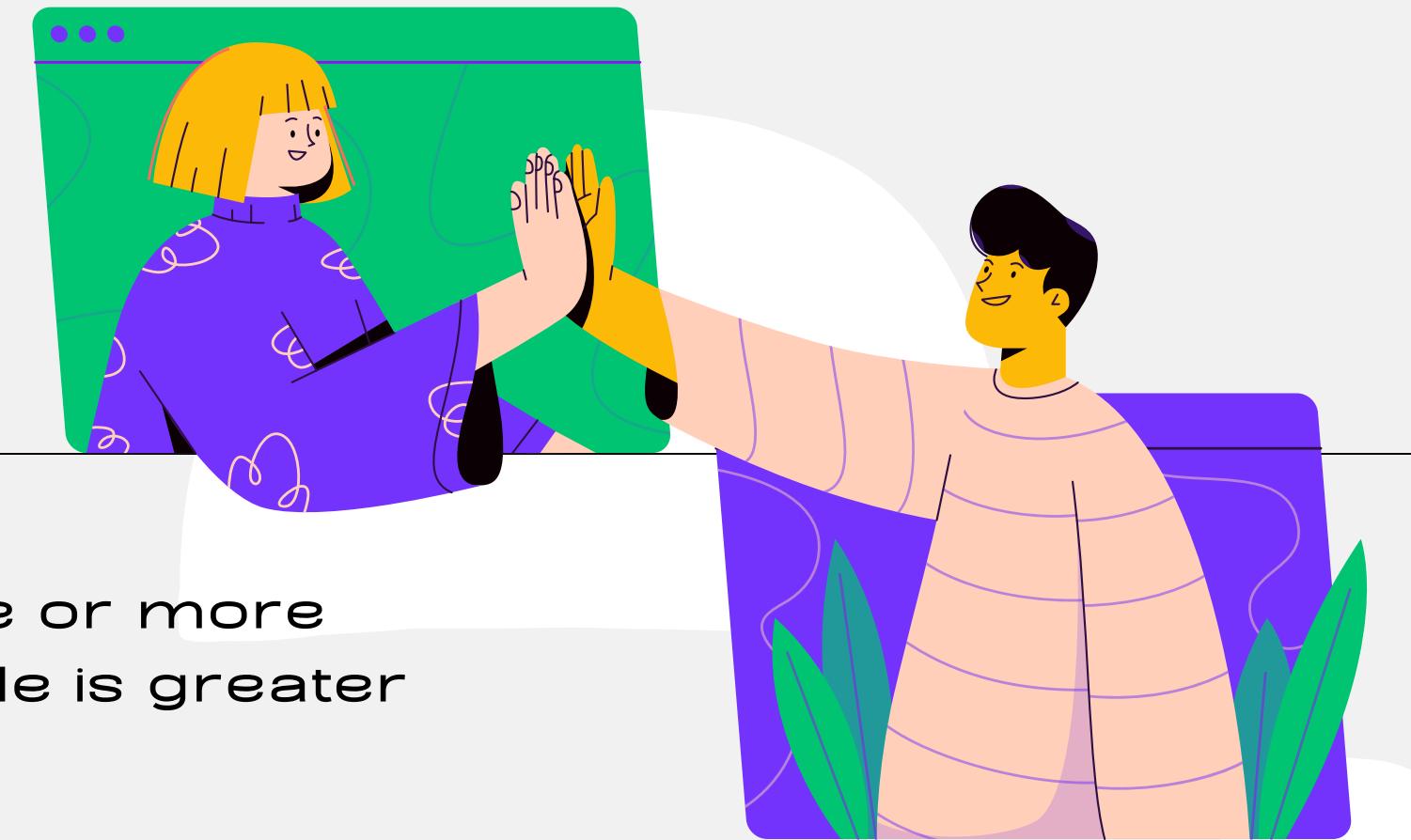
```
SELECT  
    (SELECT DISTINCT  
        Salary  
    FROM  
        Employee  
    ORDER BY Salary DESC  
    LIMIT 1 OFFSET 1) AS  
SecondHighestSalary  
;
```



Source:  
Leetcode

# SQL Advanced

## Facebook



Write an SQL query to display the records with three or more rows with consecutive id's, and the number of people is greater than or equal to 100 for each.

Return the result table ordered by visit\_date in ascending order.

Table: Stadium

Column Name	Type
<code>id</code>	<code>int</code>
<code>visit_date</code>	<code>date</code>
<code>people</code>	<code>int</code>

visit\_date is the primary key for this table.  
Each row of this table contains the visit date  
and visit id to the stadium with the number of  
people during the visit.  
No two rows will have the same visit\_date, and  
as the id increases, the dates increase as well.

Source:  
Leetcode

# SQL Advanced

## Facebook



### Solution

```
select distinct t1.*  
from stadium t1, stadium t2, stadium t3  
where t1.people >= 100 and t2.people >= 100 and t3.people >=  
100  
;
```

Source:  
Leetcode

# Python Basics

## Deep copy:

- It constructs a new compound object and then, recursively, inserts copies into it of the objects found in the original.
- It makes the reference to an object and the new object that is pointed by some other object gets stored.
- The changes made in the original copy won't affect any other copy that uses the object.
- It makes execution of the program slower due to making certain copies for each object that is being called.

## Shallow copy:

- It constructs a new compound object and then [to the extent possible] inserts references into it to the objects found in the original.
- Shallow copy is used to copy the reference pointers just like it copies the values.
- These references point to the original objects and the changes made in any member of the class will also affect the original copy of it.
- It allows faster execution of the program and it depends on the size of the data that is used.

# Python Basics

**Map** → Utility function, maps a collection to another collection object based on certain functionality.

`map(function, iterable object)`

For example: If we have list of people like:

```
firstname = ["Ram", "Shyam", "Vinay", "Gopal"]
```

- Map the list to obtain the names in upper case
- `list[map[lambda x:x.upper(), firstname]]`

**Filter** → Similar function, but it requires the function to look for a condition and then returns only those elements from the collection that satisfies the condition.

**Reduce** → An operation that breaks down the entire process into pair-wise operations and uses the result from each operation, with the successive element.

# Python Basics

## what is the Lambda Function?

- Lambda functions are an anonymous or nameless function.
- These functions are called anonymous because they are not declared in the standard manner by using the `def` keyword. It doesn't require the `return` keyword as well. These are implicit in the function.
- The function can have any number of parameters but can have just one statement and return just one value in the form of an expression. They cannot contain commands or multiple expressions.
- An anonymous function cannot be a direct call to `print` because lambda requires an expression.
- Lambda functions have their own local namespace and cannot access variables other than those in their parameter list and those in the global namespace.

Example: `x = lambda i,j: i*j`

```
print(x[2,3])
```

Output: 6

# Python Basics

## Lists vs Tuple

1. Items surrounded in square brackets []
2. Lists are mutable in nature
3. There are more than 40 available methods in Lists.
4. If content is not fixed, and keeps on changing then we should go for lists.
5. List objects cannot be used as keys for dictionaries because keys should be Hash table and immutable.

1. Items surrounded in round brackets ()
2. Tuples are immutable in nature
3. There are almost 30-35 available methods in Tuples.
4. If content is fixed, and never changes then we should go for Tuples.
5. Tuple objects can be used as keys for dictionaries because keys should be Hash table and immutable.

# Python Basics

## What is the difference between pass, continue & break

- Pass: It is used when you need some block of code syntactically, but you want to skip its execution. This is basically a null operation. Nothing happens when this is executed.
- Continue: It allows to skip some part of a loop when some specific condition is met and the control is transferred to the beginning of the loop. The loop does not terminate but continues on with the next iteration.
- Break: It allows the loop to terminate when some condition is met and the control of the program flows to the statement immediately after the body of the loop. In case, the break statement is inside a nested loop [the loop inside another loop], then the break statement will terminate the innermost loop.

# Python Basics

## what does enumerate() function do?

- The enumerate() function assigns an index to each item in an iterable object that can be used to reference the item later. It makes it easier to keep track of the content of an iterable object.
- Example:

```
list2 = ["apple","ball","cat"]
```

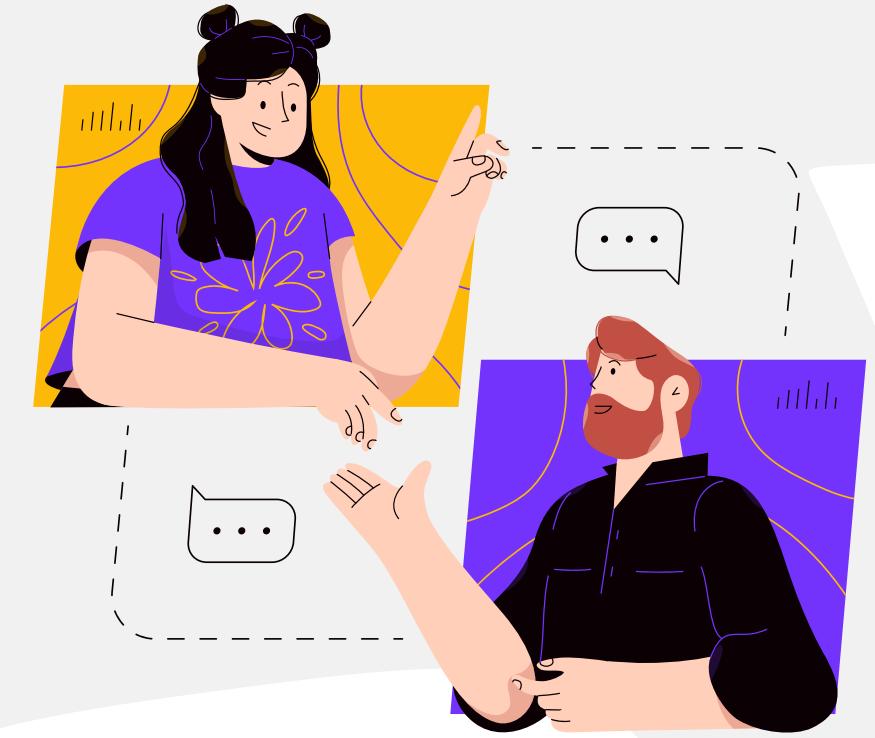
```
e1 = enumerate(list2)
```

```
print(e1)
```

**Output:** [[0, 'apple'], [1, ball'], [2, 'cat']]

# Python Assignment

## Data



WeekID	Week	Area	LocationDesc	Category	ColorDesc	SubCategory	StyleID	FabricID	ColorID	Sum of SalesEUR	S	Dnt	Sum of OnHandEUR
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	CARMINE+FLAM+BLACK	Handbags 1	3705880	ACMEO2T	11117621			5	8069.372
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	CARMINE+N.L.RED+BLAC	Handbags 1	3705880	ACMEO2Y	11117621			3	4841.394
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878878	A1AEXB0	11636615	1613.798	1	3	4841.394
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	DEEP IVORY	Handbags 1	3878878	A1AEXB0	11636615			2	3227.978
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	LACQURED RED	Handbags 1	3878878	A1AEXBT	11636615			6	9682.788
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878878	A1HEBOT	11636615			7	11296.204
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878878	A1HEBOY	11636615			9	14524.183
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878878	ACMEOOT	11636615			9	14524.182
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	DARK PURPLE	Handbags 1	3878878	ACMEOOT	11636615			6	9682.788
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	DARK ROSE	Handbags 1	3878878	ACMEOOT	11636615			4	6454.81
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	OFF WHITE	Handbags 1	3878878	ACMEOOT	11636615			5	8068.226
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878878	ACMEOOY	11636615			10	16137.598
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	LUST RED	Handbags 1	3878878	ACMEOOY	11636615			2	3227.978
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878892	ACMEOOT	11636657			4	7219.12
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3878892	ACMEOOY	11636657	1804.78	1	4	7219.12
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3909769	A1AEXB0	11729288			10	15183.07
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	DEEP IVORY	Handbags 1	3909769	A1AEXB0	11729288			6	9109.842
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3909769	A1BE110	11729288			11	18801.797
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3909769	A1BE111	11729288	1709.289	1	6	10255.734
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3909769	ACMEOOT	11729288			5	7591.535
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	OFF WHITE	Handbags 1	3909769	ACMEOOT	11729288			8	12146.456
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	BLACK	Handbags 1	3909769	ACMEOOY	11729288			4	6073.228
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	ANEMONE	Handbags 1	3909769	ACMEOOY	11729288			6	9109.842
1301	Wk. 1 January 2019	USA	New York Store	Handbags Women	LUST RED	Handbags 1	3909769	ACMEOOY	11729288			3	4554.921

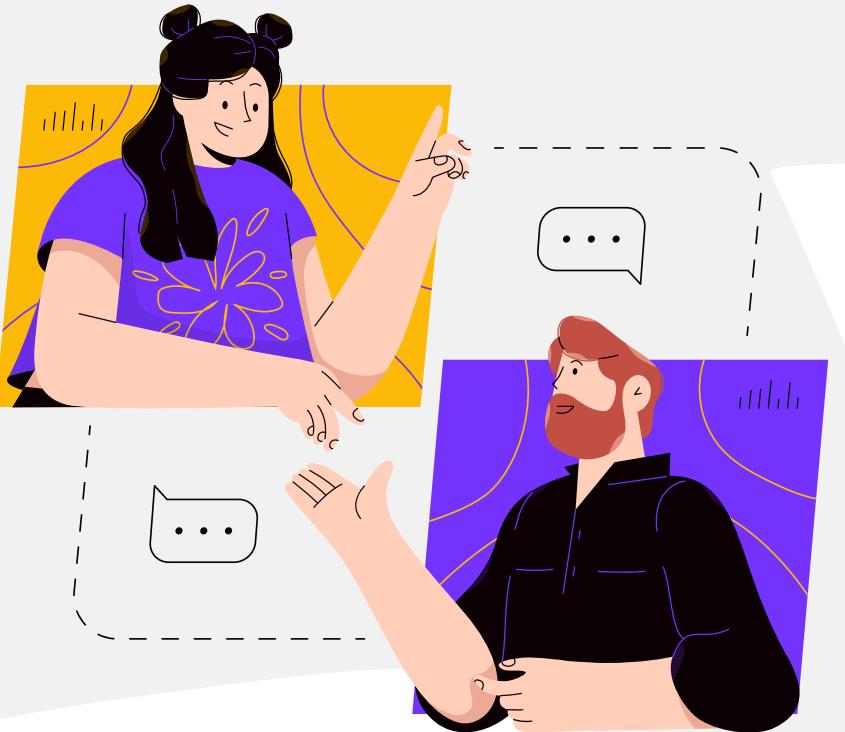
# Questions

## Pandas manipulation

The purpose of this section is to evaluate your ability to use pandas and to manipulate data.

In the excel file DataTest.xlsx, you have sales and stock (onhand) data. A product is defined by the StyleID, FabricID and ColorID.

- 1) read the excel file and create two columns : Year and Month
- 2) replace negative sales/onhand by 0
- 3) Create a columns which tell if a product is sell only in USA, only in China or both
- 4) Per category Retrieve the top 3 products (common to both regions) (optional)
- 5) What is the top 3 Color Description per SubCategory during the first half of 2020 ?
- 6) Create a columns which return
  - high sales :if the Sales qty are > 10 and the category is Handbags Women or the subcategory is Shoes 1
  - high sales :if the sales qty are >20 and the subcategory is Shoes 2
  - normal sales :if it's not high sales and the sales qty >2
  - low sales :else
- 7) Show the evolution of the average price per subcategory
- 8) Create a column Adjusted Sales which is the sales + the missing sales (highlight the missed opportunity and caculate the missing sales)
- 9) Create a recommendation of the 10 products per category with a high potential (optional)
- 10) Knowing the time of production is 3 months and I can't transfer stock between regions, what should I reorder for the next 6 months (per product per region)? (optional) (if you can comment your strategy, what are the strength and the weakness of the output)



For Solutions:  
[Click here](#)

# Statistical Questions

what's your knowledge on statistics, and how you have used it in your work?

If possible describe how you've used stats to solve a problem.

**Example:** You can tell them how you imputed the null values, what statistical strategies you took, you can explain them how to perform EDA, categorical vs numerical analysis etc [Don't mug up, try to answer as openly from personal experiences as possible]

which step of a data analysis project did you enjoy the most?

It's situational based, and depends on personal interests.

**Example:** I have enjoyed performing the EDA steps, so that I can know more about the data, have performed categorical vs numerical analysis, performed correlation on numerical variables and so on..

# Dashboard Building Questions

## what's your experience in creating dashboards?

Try explaining the interviewer how dashboards are important, how can we capture the KPIs and metrics, and make it look visually appealing and track business goals. You can then explain about one of the tools that you have used, it could be Power BI, MS Excel, Tableau, or anything else, and if possible, talk about some of the features of the BI tool you used.

## Explain some of the important charts, and where

There are many basic charts you should be aware of, such as area charts, bar charts, column charts, doughnut charts, pie charts, gauge charts etc.

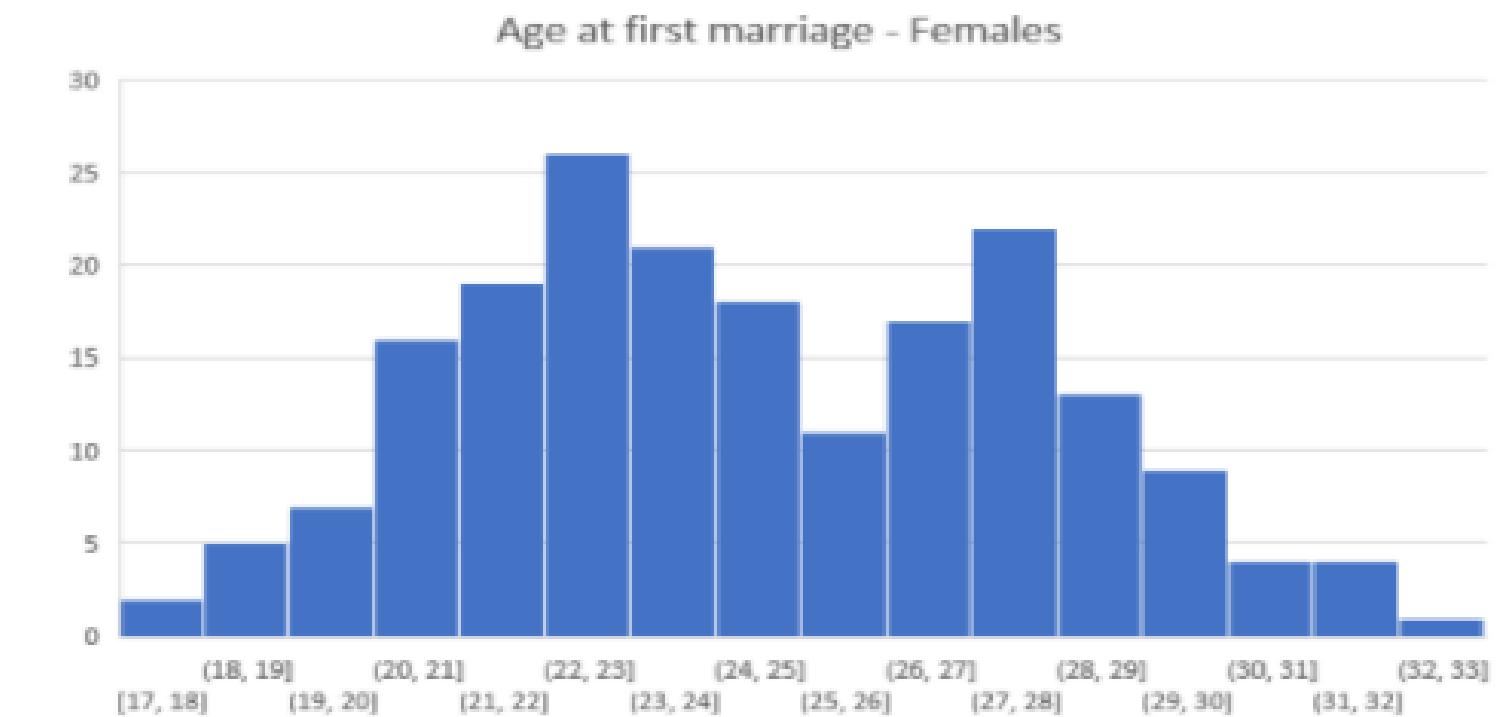
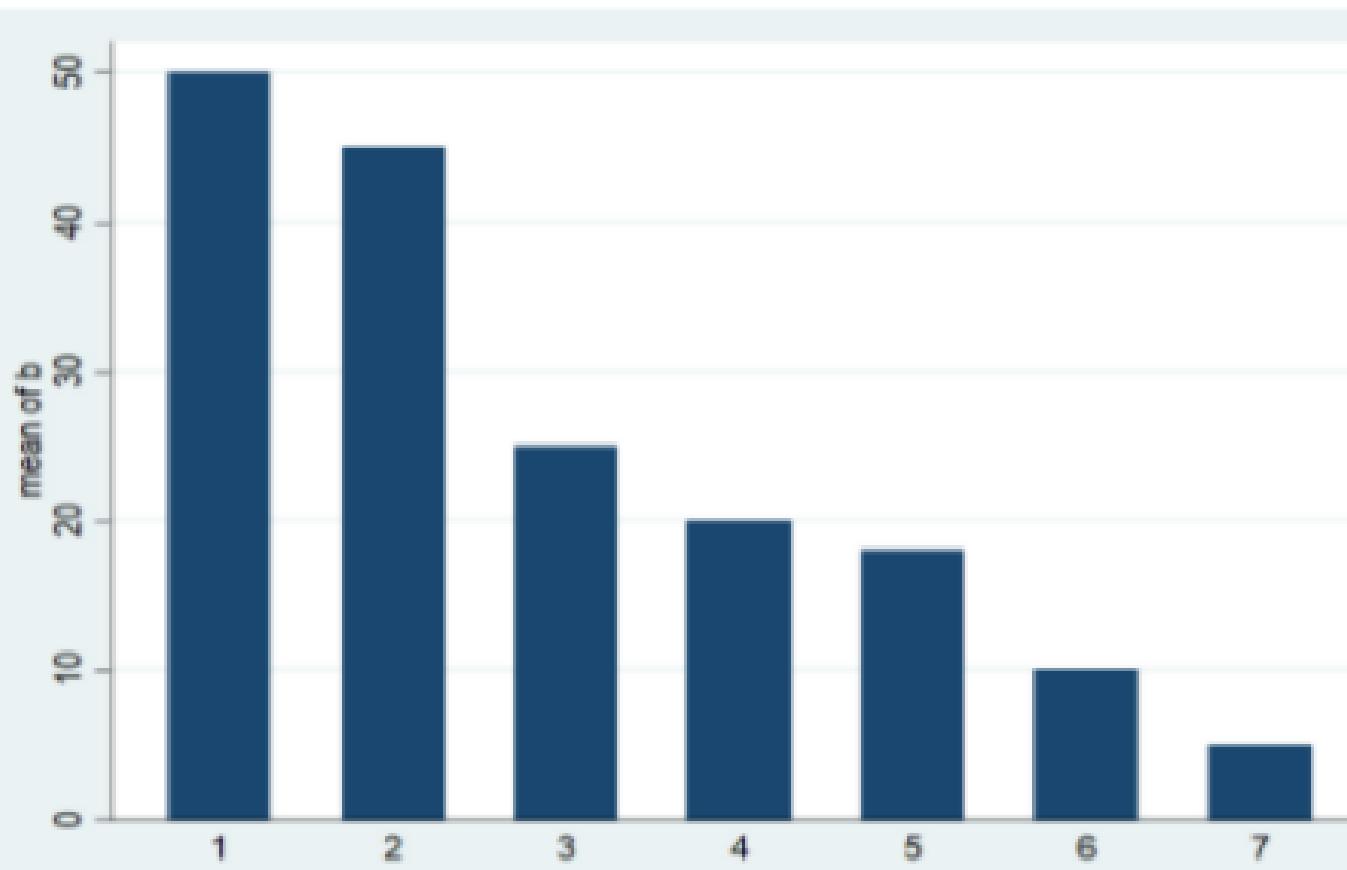
Also explain, which chart to be used for which particular scenario

# Data visualization Charts

## Important Charts for Visualisation

### Histogram

Histogram represent the frequency distribution of the data



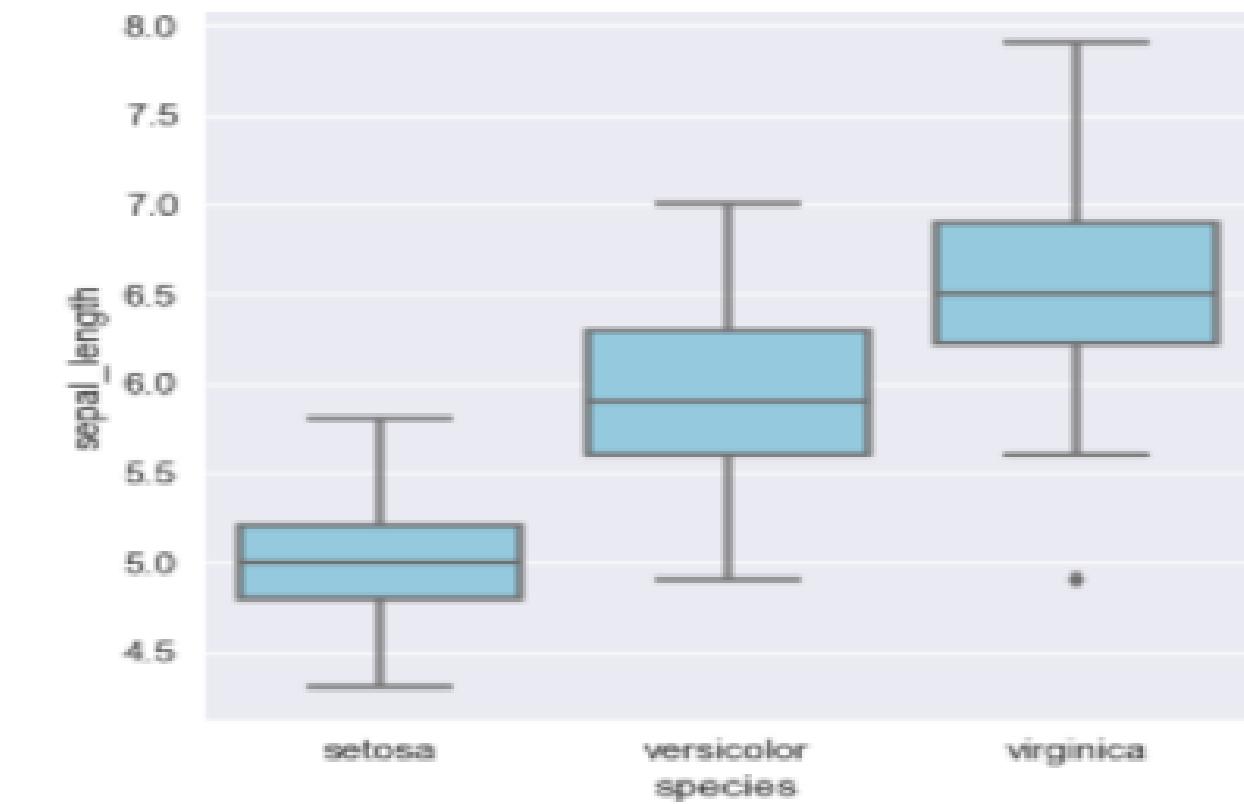
### Bar Chart

Bar graph represent the total observation in the data for a particular category.

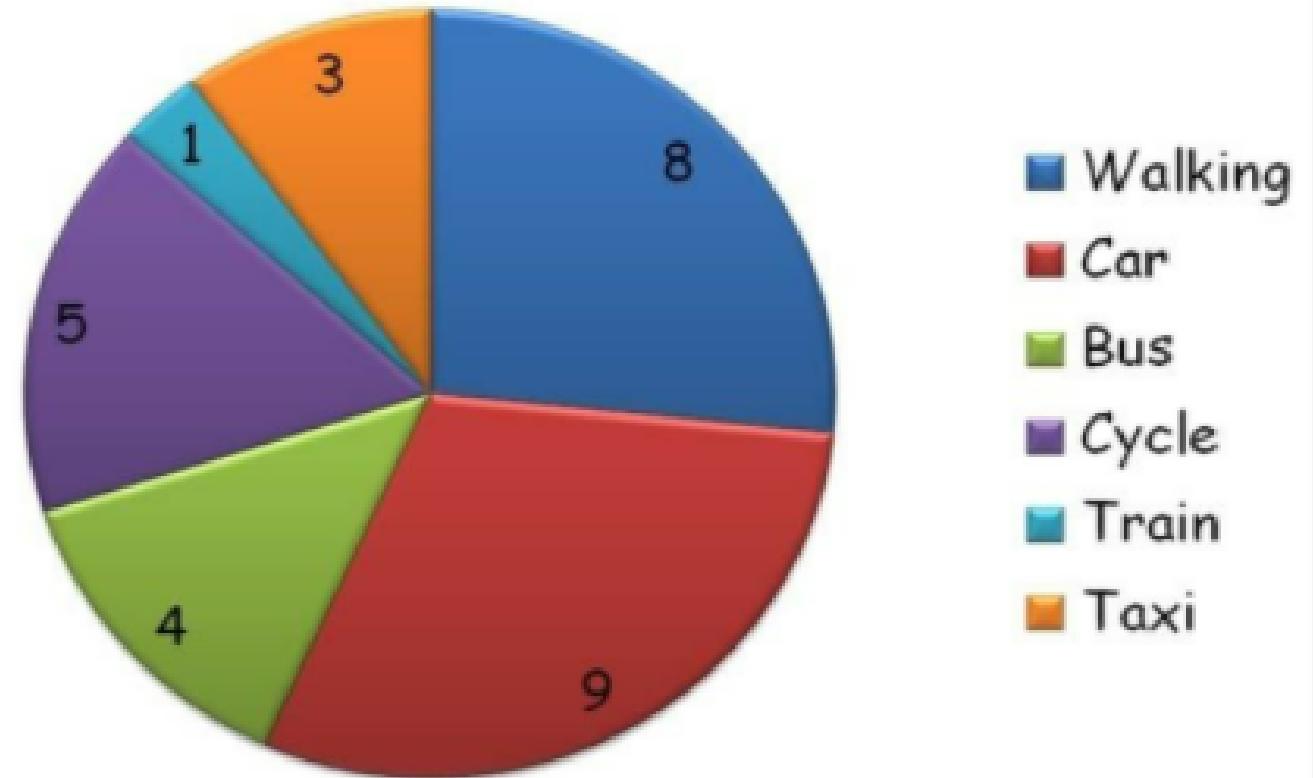
# Data visualization Charts

## Box Plot

Boxplot display the distribution of the data based on five number summary(minimum, first quartile, median, third quartile, maximum)



Methods of Travelling to School



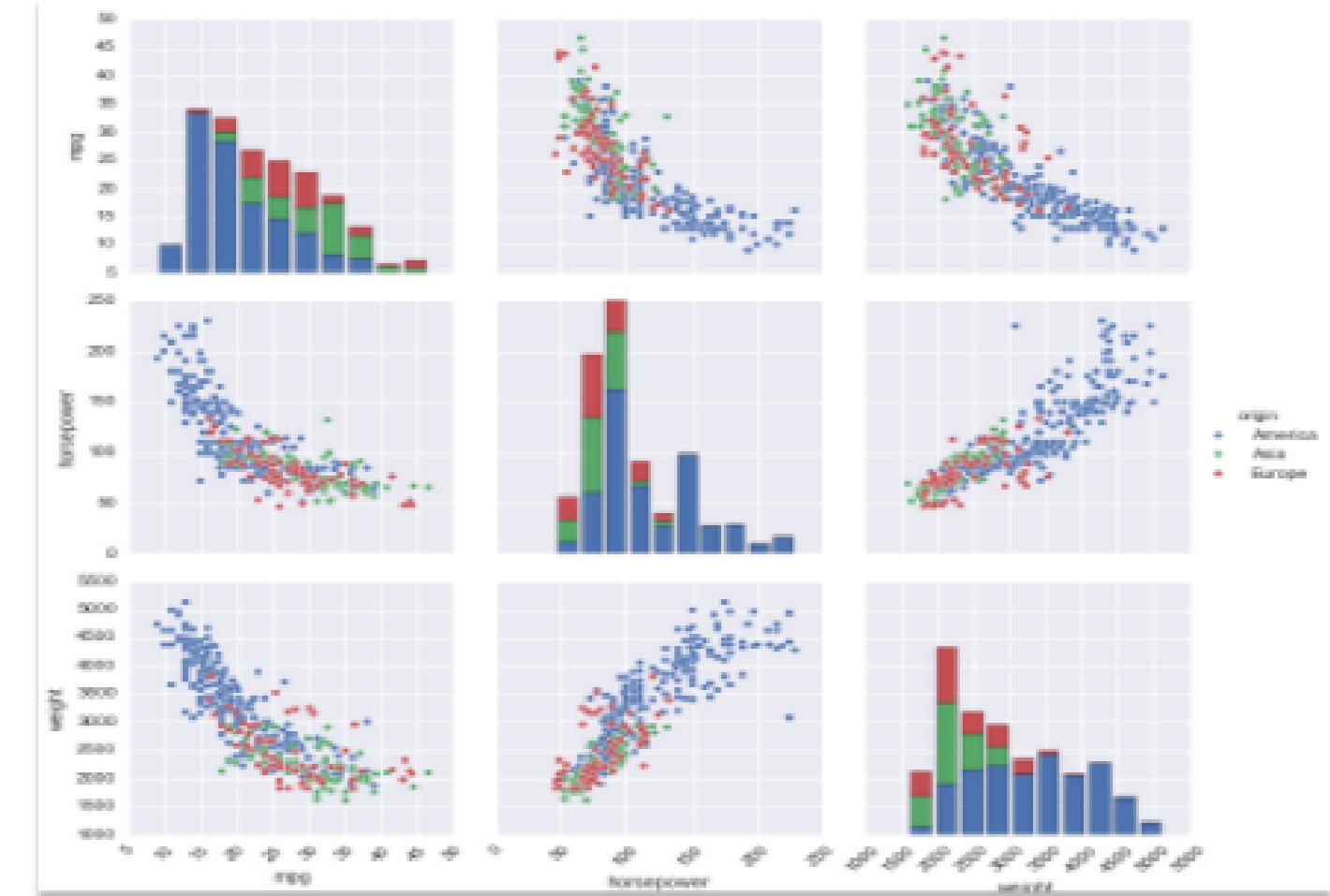
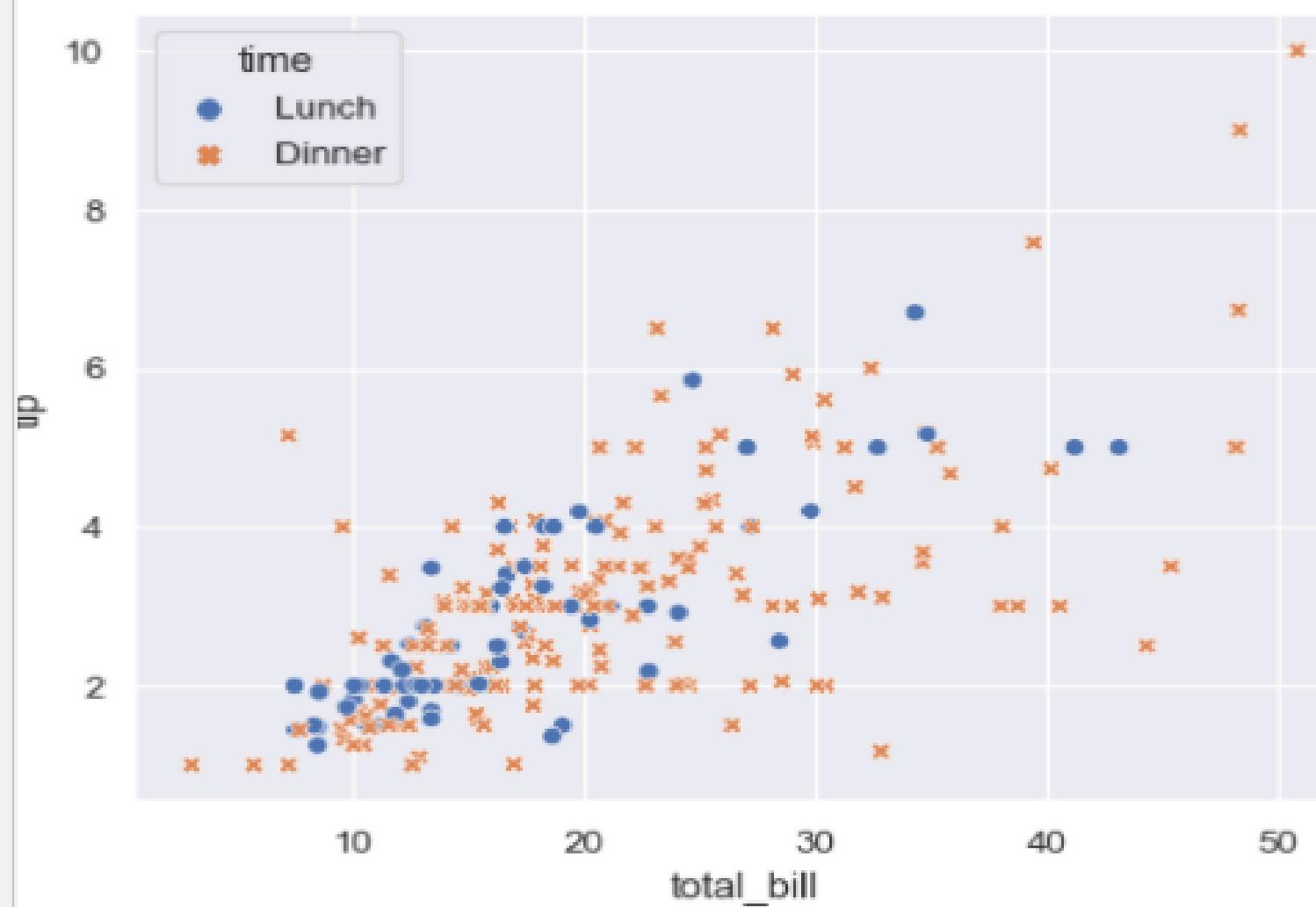
## Pie Chart

Pie chart represent the percentage of the data by each category.

# Data visualization Charts

## Pair-plot

Pair plot show the bivariate distribution of the datasets. It show the pairwise relationship between the variable



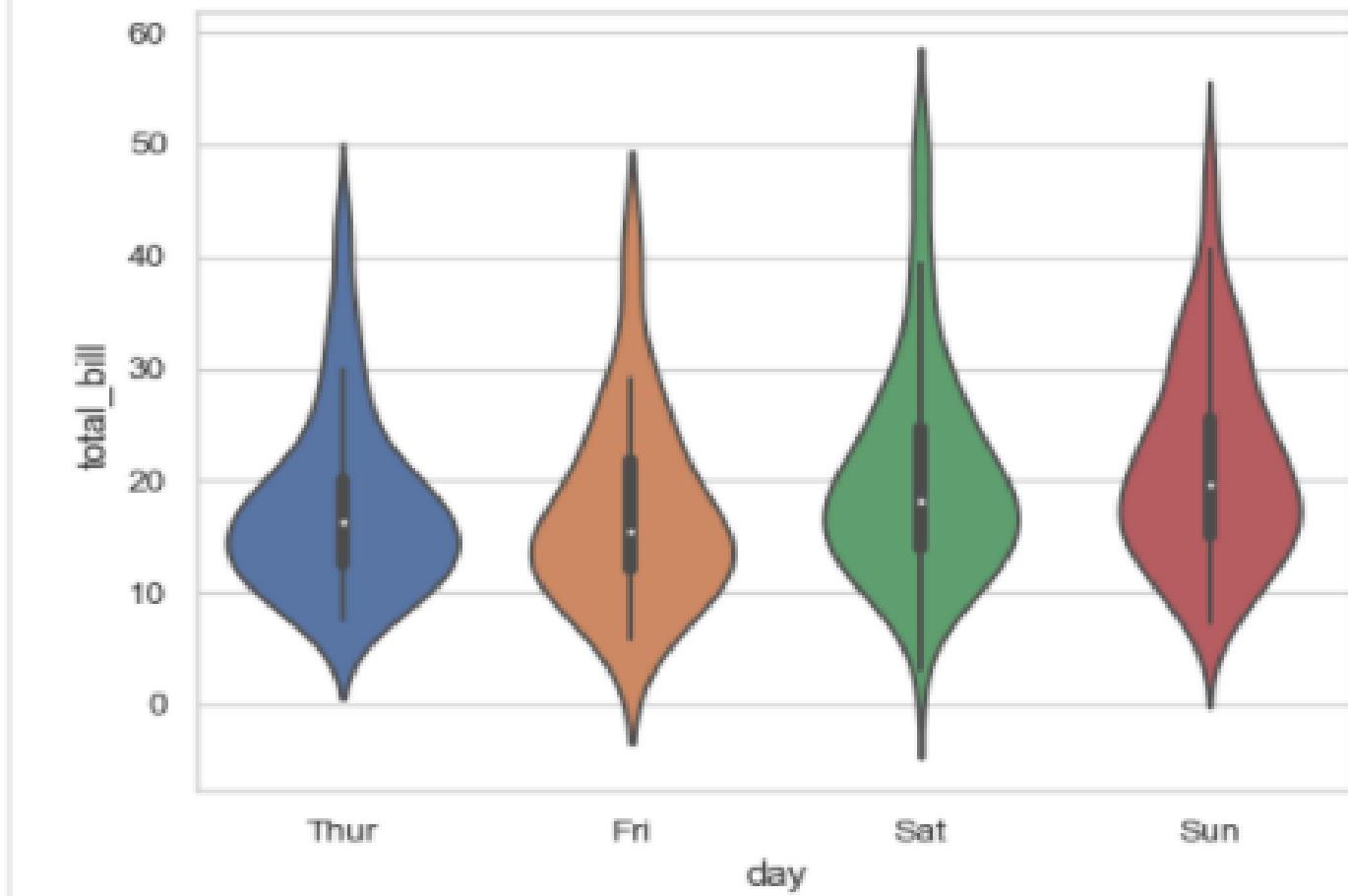
## Scatter Plot

Scatter plot represent the relationship between two numerical variable. It show the correlation between two variable.

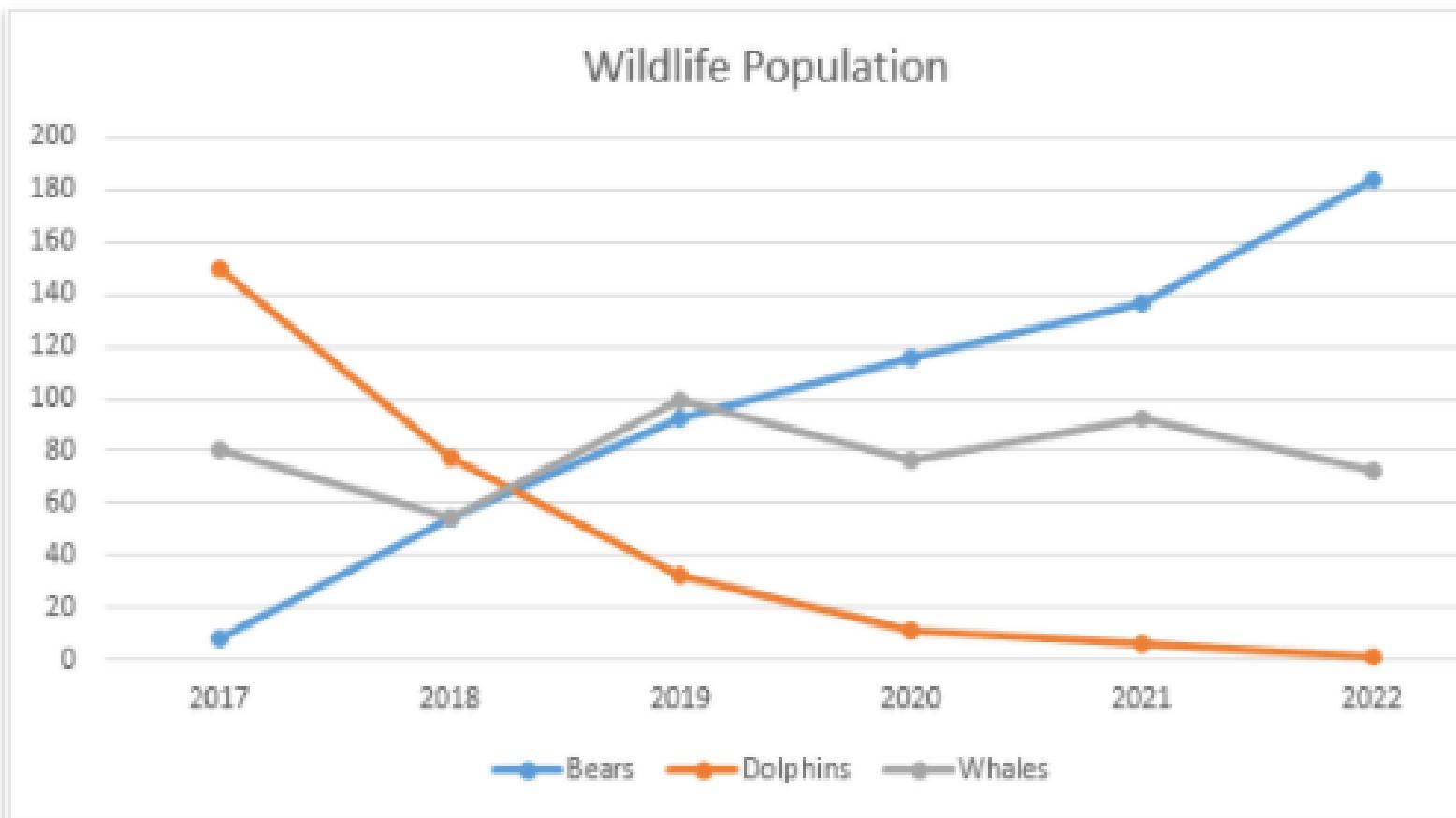
# Data visualization Charts

## Violin Plot

Violin chart are used to plot numeric data



Wildlife Population



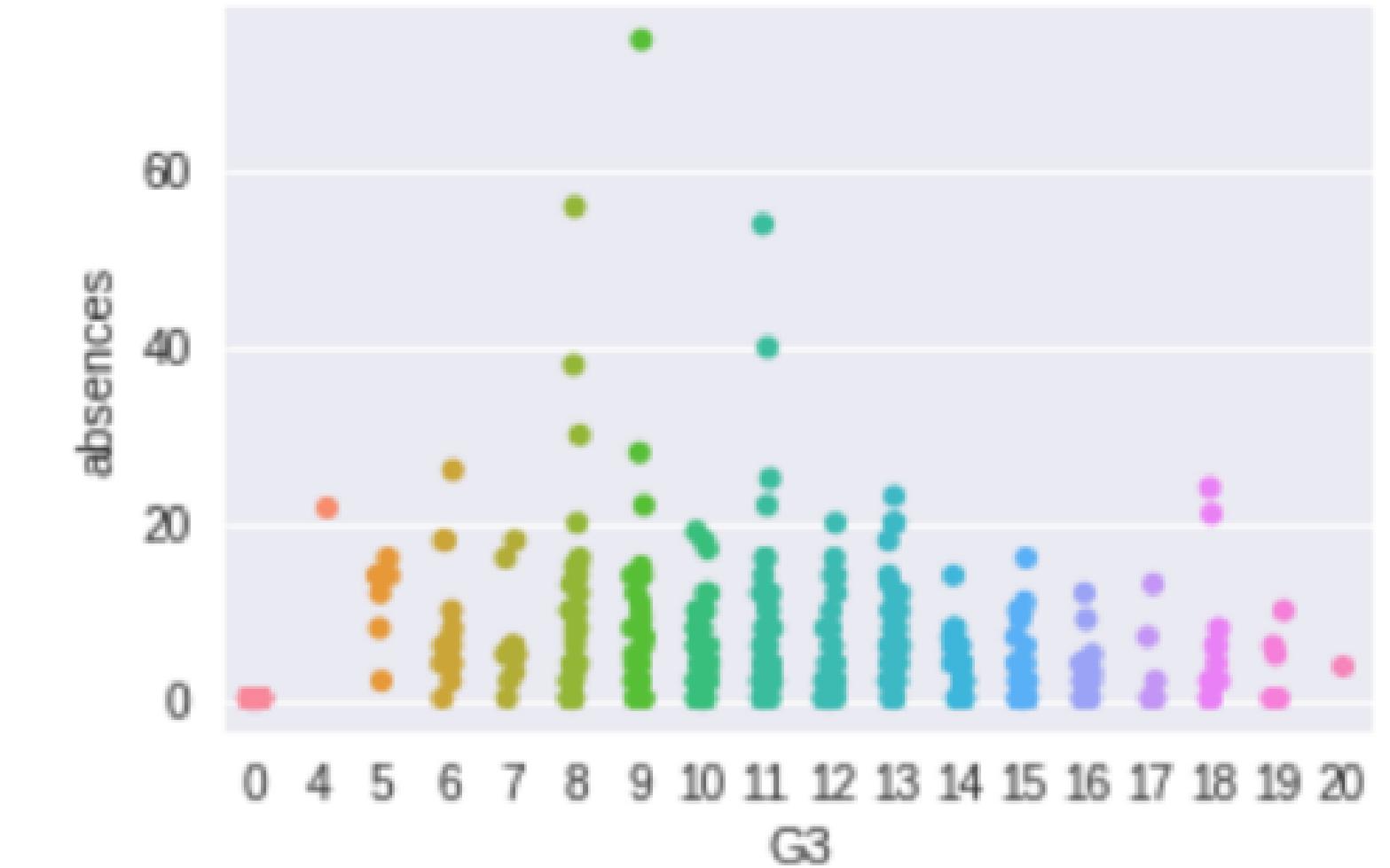
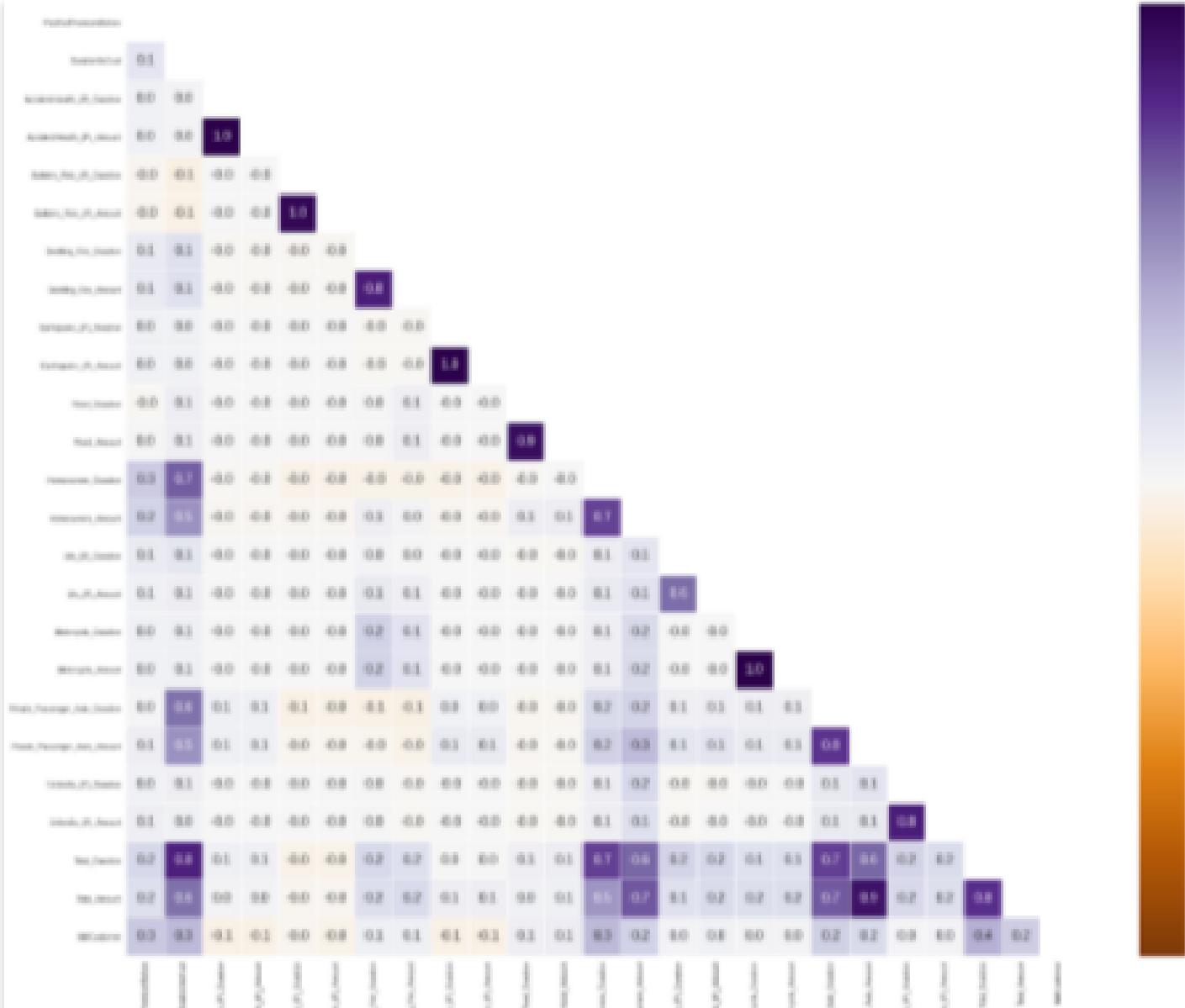
## Line Chart

Line chart are used to track change over line and short period of time. Line chart are used in time series data.

# Data visualization Charts

## Strip Plot

Strip plots are used to draw a scatter plot for on the categorical basis



## Heatmaps

A heat map is data analysis software that uses color the way a bar graph uses height and width.

# Power BI Questions

## FILTER FUNCTION

Let say, you want to find the total count of customers present in Mumbai

```
Count_Mumbai =  
CALCULATE(COUNT('table'[CUST_ID]),  
FILTER('table', 'table'[CITY] = "MUMBAI"))
```

FILTER acts as the WHERE clause



How would you create trailing X month metrics via DAX against a non-standard calendar?

The solution will involve:

1. CALCULATE function to control [take over] filter context of measures.
2. ALL to remove existing filters on the date dimension.
3. FILTER to identify which rows of the date dimension to use.

Alternatively, CONTAINS may be used:

- CALCULATE(FILTER(ALL(['DATE']),.....))

# Power BI Questions

**Can we have more than one active relationship between two tables in data model of power pivot?**

No, we cannot have more than one active relationship between two tables. However, can have more than one relationship between two tables but there will be only one active relationship and many inactive relationship. The dotted lines are inactive and continuous line are active.

**Explain when Do You Use `Sumx[]` Instead of `sum[]`?**

when the expressions to `SUM[]` consists of anything else than a column name. Typically when you want to add or multiply the values in different columns:

`SUMX[Orderline, Orderline[quantity], Orderline[price]]`

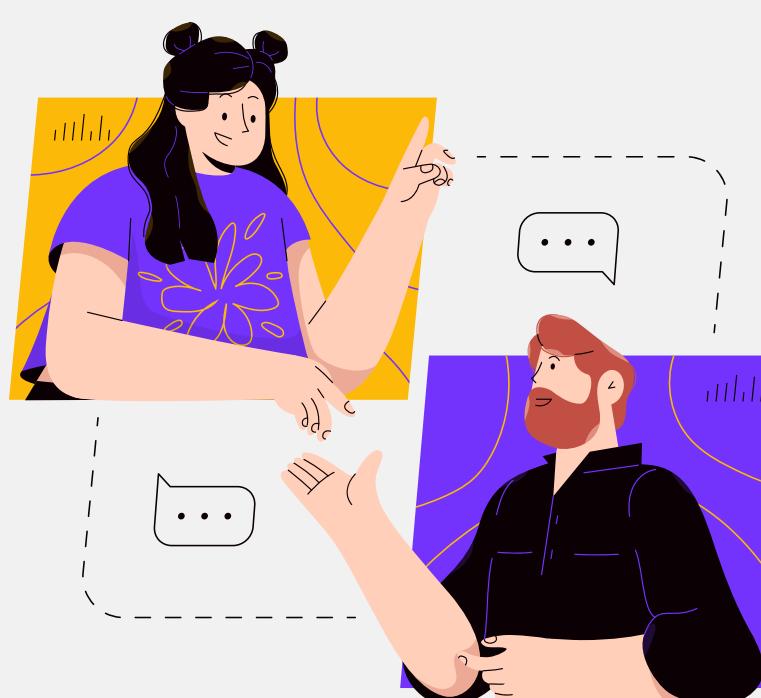
`SUMX[]` first creates a row context over the Sales table (see 1 above). It then iterates through this table one row at a time. `SUM[]` is optimized for reducing over column segments and is as such not an iterator.

# Power BI Questions

## Name Any 3 Most Useful Text Functions In Dax?

The text functions in DAX include the following:

- CONCATENATE
- REPLACE
- SEARCH
- UPPER
- FIXED



## what Is The Difference Between DISTINCT[] And VALUES[] In Dax?

Both count the distinct values, but VALUES[] also counts a possible implicit virtual empty row because of non matching values in a child table. This is usually in a dimension table.

## which Function Should You Use Rather Than COUNTROWS(DISTINCT[])?

DISTINCTCOUNT()

# Power BI Questions

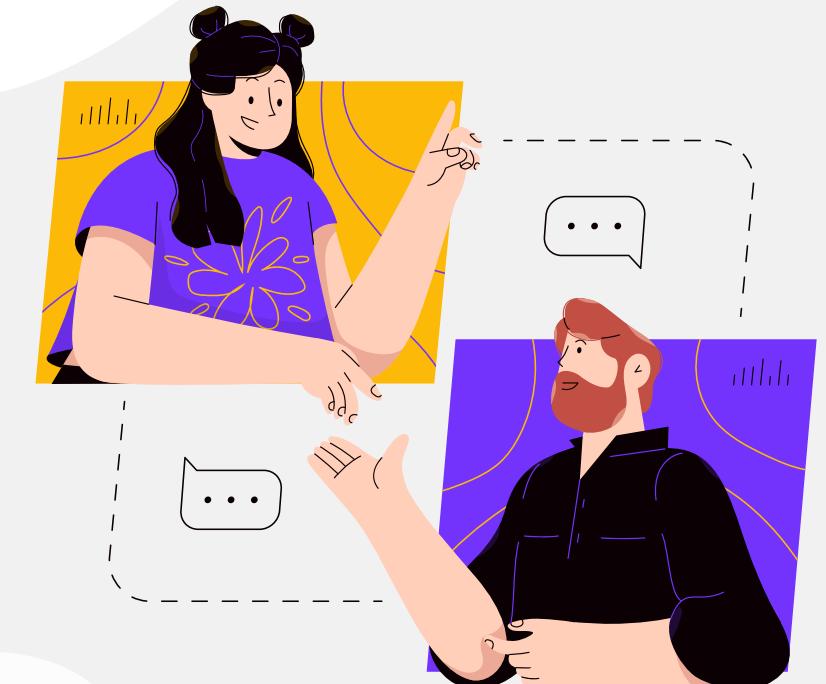
## what is Power Query

Power query is a ETL Tool used to shape, clean and transform data using intuitive interfaces without having to use coding. It helps the user to:

- Import Data from wide range of sources from files, databases, big data, social media data, etc.
- Join and append data from multiple data sources.
  - Shape data as per requirement by removing and adding data.

## Common Power Query Editor Transforms?

Changing Data Types, Filtering Rows, Choosing/Removing Columns, Grouping, Splitting a column into multiple columns, Adding new Columns ,etc.

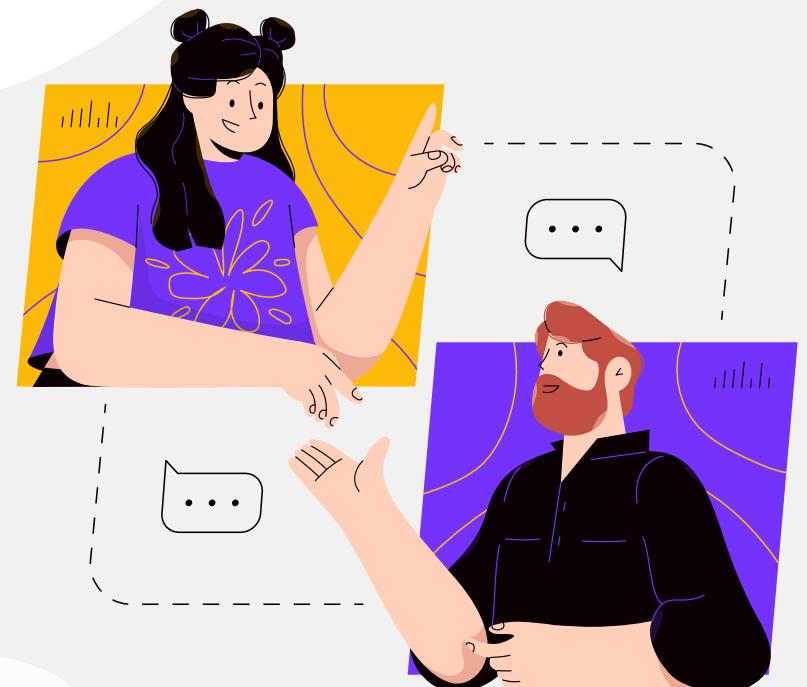


# Basic ML Questions

## How to handle missing values & null values?

One of the following methods can be applied while handling missing values, depending on the nature of the data and data type we are working with

- Drop Rows with more than 50% missing values
- Impute missing values with central tendency measures such as mean or median
- For categorical attributes, impute missing values with the most frequent category [mode]
- Predicting the missing values with regression or classification models
- Using machine algorithms such as K-NN that supports missing values while making a prediction



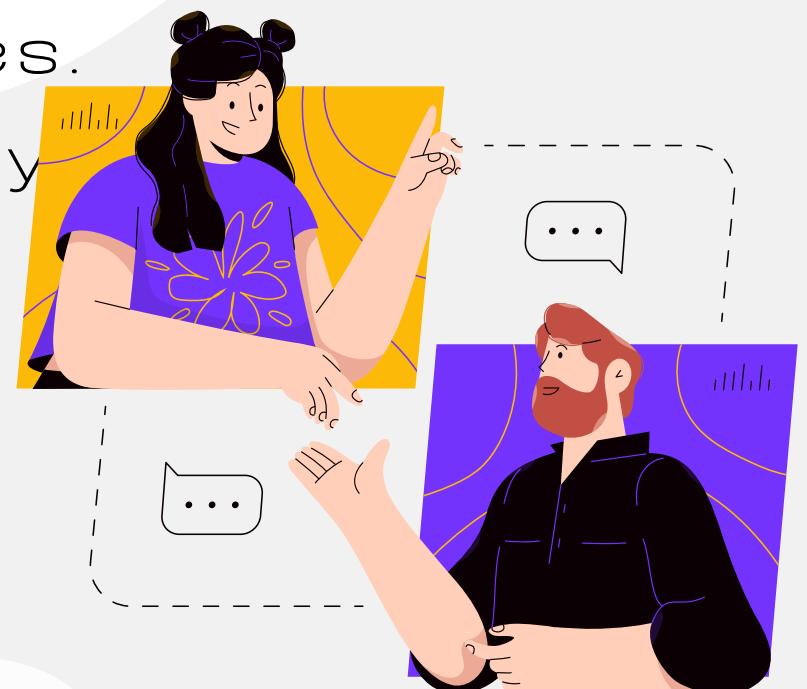
# Basic ML Concepts

## Explain univariate, bivariate & multivariate analysis

**Univariate analysis:** Univariate is a form of data analysis where a single variable is analyzed to describe and find patterns that exist within it. It is the simplest form of data analysis as it doesn't deal with causes or relationships.

**Bivariate Analysis:** Bivariate analysis measures the correlation between two variables. This technique is used by researchers when they aim to draw comparisons between two variables.

**Multivariate Analysis:** Multivariate analysis is used to study complex data sets. In this form of analysis, a dependent variable is represented in terms of several independent variables observations available to establish such a relationship.



# Basic ML Concepts

**Different types of Machine learning techniques?**

Supervised Learning: Dealing with the labeled data.

Unsupervised Learning: Unlabeled data

Reinforcement Learning: In reinforcement learning, decisions are made by the system based on the feedback it receives for its actions. In this approach, the algorithm learns from its mistakes and improvises to return better results, over time.

**Explain any of the Machine Learning project you did**

Talk about any project you did in past.

Always explain 4 major things:

1. Business Understanding: what the use case is all about, what business problem is it solving
2. Exploratory Data Analysis: Tell some important insights that you got by performing any EDA steps
3. Model Building: Talk about the algorithms you used.
4. Model Deployment

**More  
advanced  
questions.**

Stay tuned!!

