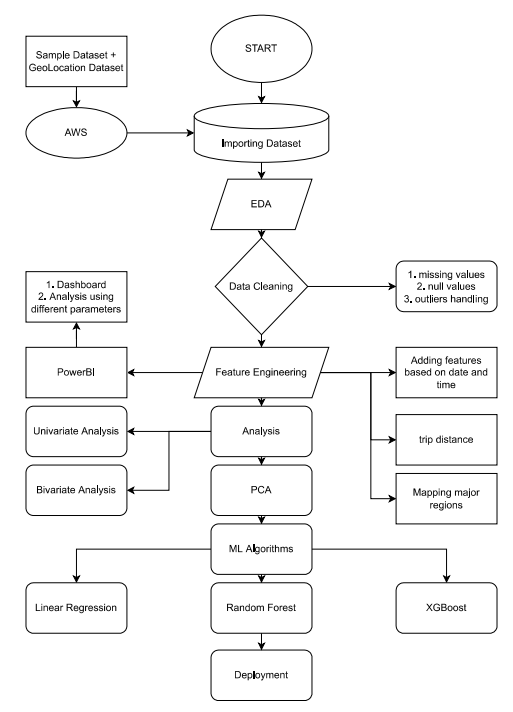
1. Can you explain your C-DAC project? Draw block diagram as appropriate.



2. What's new in your project?

1. Analysed data based on different situations like rush hour, mid night trips, snowfall, airport trip
2. Analysed data based on major areas of NYC
3. Intending to implement recommendation system on top of fare prediction model in coming days

3. Explain why did you select this technology & framework for this project?

1. Python
   1. Suitable for data manipulation
2. Machine Learning
   1. Used models like :
      1. Linear regression
      2. XGBoost
      3. Random Forest
   2. Techniques:
      1. Principal component analaysis
3. Flask
   1. Used for server deployment
   2. Used with HTML and CSS
4. PowerBI
   1. Used for creating interactive Dashboards
   2. Also used for presentation
5. AWS
   1. Used for data storage
   2. Used for deploying model on cloud
6. Explain n-tier architecture of your project?
7. It can be said as 2-tier architecture
8. As it consists of logic tier and presentation tier
9. But it doesnt contain data tier

5.Which advanced features have you used in your project ?

1. PowerBI

a. custom visualization

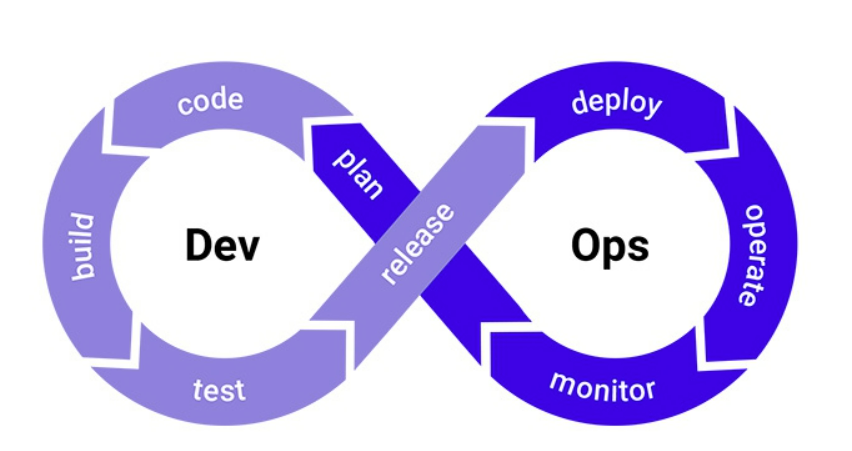
b. data connection and shaping

2. AWS

a. S3 storage

b. AWS EC2 instance

6. What was your role in your project and explain what you did in it?

1. This project was completed by end-to-end alone.. from data collection to deployment on cloud..
2. Which software development methodology(model/architecture) you have used in project? Explain its process.
3. DevOps deployment method…
4. 
5. More focused on improving time to market
6. lowering the failure rate of new releases
7. minimizing the disruptions while maximizing reliability
8. aims to automate continuous deployment

8. How will you deploy your project (on web server/client machine)?

1. I have created the lightweight web server using the flask framework with the support of html and css

Then created the AWS EC2 instance on which I uploaded the required codes to run the server on cloud.

And then executed the program which activated the server online through cloud

9. Which Design pattern are used in your project?

10. What are the limitations of your project?

1. due to lack of fuel price dataset, unable to gauge the exact profit

2. incomplete geolocation mapping of the points

3. too small sample taken from the original dataset

11. What are the difficulties you have faced during this project and How you have overcome it?

1. finding the longitude and latitude of zones based on the address. After going through multiple options, I decided to use the geopy library which almost returns the original values

2. finding the distance between two locations was another challenge, instead of giving the actual distance it was giving the distance as straight line between two points.

3. I searched online and used two methods to find the distance between location points, and took the mean of it for further use, to reduce the calculation

12. How will you improve the performance of your project? (memory related and response time) ?

13. Which database is used in your Project? Why? Explain database design.

14. Explain security of your project?

15. Mention sources of your dataset?   
1. NYC taxi limunoiuse commission website is the one where I could find the original raw dataset

16. What is application of your project?

To predict the revenue generated by a single taxi taking pickup and drop off location of vendor as a input

17. Name some feature extraction techniques used for dimensionality reduction in your project?

1. having more than 50 features in a dataset, I used the technique to find the most impactful feature on the machine learning model, for this I used the principal component analysis

2. after this I found the features like fare\_amount, tip\_amount and total\_amount to be the independent features

3. And pickup\_location, dropoff\_location and vendorid as the dependent features

18. How much data you have allocated training, validation and testing in your project?

1. I have allocated 80% data for training and 20% data for testing

19. Write map-reduce logic of your project?

20. How SDLC is implemented in your project?

21. Mention which steps you followed in your Analytics project.

22. Which data cleaning and validation methods used in your project?

1. removed the negative fare and distance values

2. removed the locations which are from outside the NYC

3. removed the null and nan values and there respective records

23. Which tool you have used for data analysis? Why

1. I have used PowerBI for data analysis.

2. range of attractive visualisations

3. datasets filtration

4. informative reports

5. get data from multiple sources

6. DAX data analysis function

7. customizable dahboards