

Data Science Interview Tips & Tricks

lifecycle of data SC. Project: "PM, BA, SME, Client Architect" "Scrum Master"

① Problem/use case definition

② Data collection from one or multiple sources

DA,
DE

CSV → pd.read_csv("datafile.csv")

DB → SQLAlchemy

MongoDB → PyMongo

(AWS → Boto3)

Bucket

...

③ Data cleaning & Preprocessing.

DE
DA

→ Handle missing values

→ O/L (outliers)

→ duplicates

→ transformation, Normalization etc.

→ feature engineering

→ selection.

DA

④ EDA (Exploratory Data Analysis)

→ understand characteristics of data

descriptive stats → 5 Pt S., Info, Descript etc.

DA

→ Vis. (Histogram, Bar, Pie, Boxplots)

BA

→ Identify, trends, Correlation, aesthetics distribution etc.

⑤ Hypothesis Test

DA
MLE

Test the hypothesis) Claim / Statement
(Z, t, ANOVA, Chi Sq.) Belief.

⑤ Model Selection & Training

→ Select the right algorithm (regression, classification, clustering)

MLE
MLOPs

→ Splitting → Train, Test.

→ Normalization / Standardization (If required)

→ Train the model & Tune hyperparameters

⑥ Model Evaluate / Selection

→ Assess the performance of model.

MLE
MLOPs

Regression

- R^2
- Adj. R^2
- MSE
- RMSE etc.

Classification

- Accur -
- Precis
- Recall
- F1-Score
- etc.

Clustering

Silhouette Score

→ Compare diff models & choose best one

⑦ Deployment

→ Deploy the best model.

Integrate with the client's system.
(APIs) etc.

MLOPs
MLE

⑧ monitoring & maintenance

J8.
"MLE"

Ensure that the model remains accurate & relevant over time.

- monitor for performance & data drift
- Retrain model with new data (if necessary).

⑨ documentation / reporting

Everyone,

→ document the project processes etc.

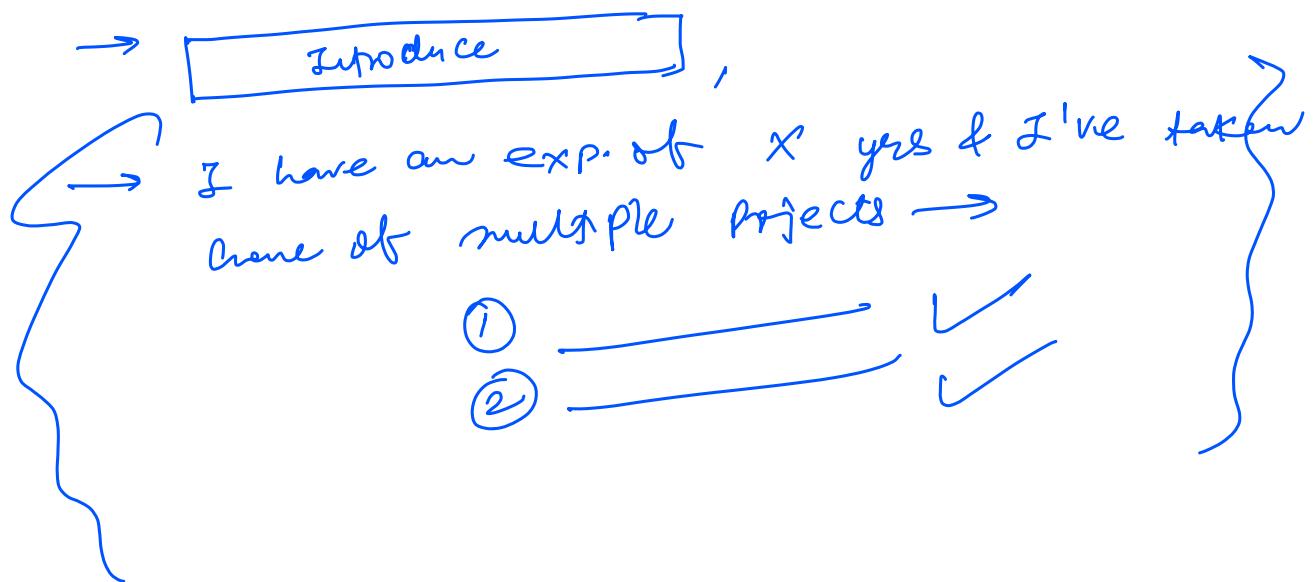
Report → Vis. Eng → Create reports to share with clients
DA

BA

Reporting E

Power BI.

① Tell me about yourself / Introduction)



② Tell me / Explain me a recent project that you've completed.

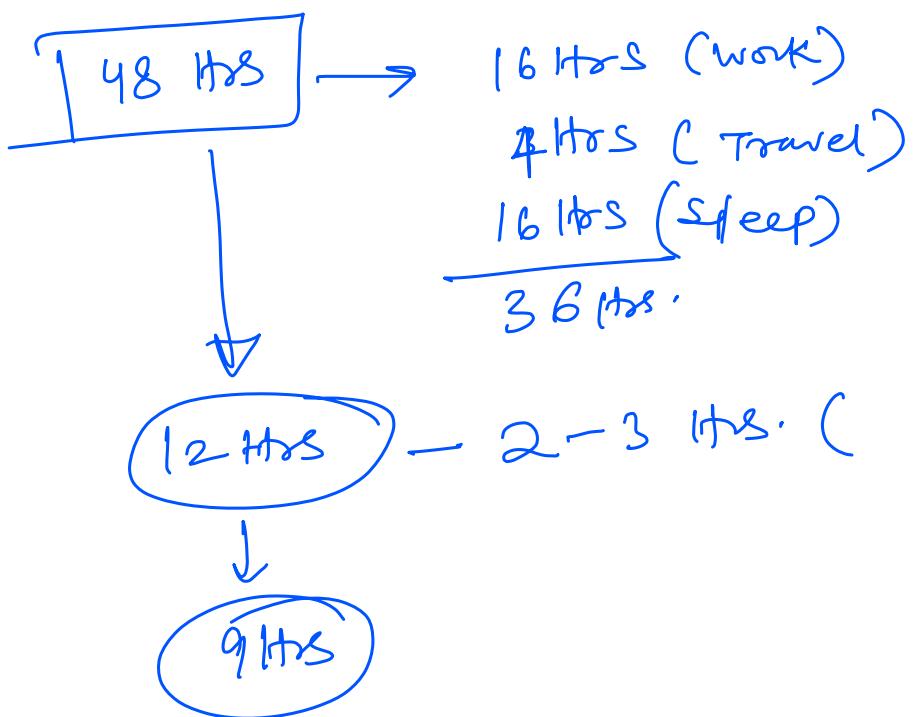
- fav. project .
- most challenging project.
- " Interesting " .

Regression / Clas

- Problem Statement
- Data columns
- size of the data
- Algo → re
- why you chose to apply an algo
& not others.
- Evaluate → R^2 , Adj R^2 ,
- Assumptions

MSE

MAE



3 hrs. Python → 1 hr (theory)
2 hrs (Practice)

1 hr. stats → descriptive,
inference
hypothesis

3 hrs ML → 2 hrs →
Theo + Pr
↓
Algo in Project

1 hr → Any other algo

1 hr. → SQL, PowerBI
Joins, windows etc.

Select, DAX,
reports.

1 Hr.

→ go through your Projects

Python:

DS → lists, dictionary, sets, tuples

Comprehension → list comprehension, dict. Comprehension

File handling → reading, writing,

Libraries → Numpy, Pandas, sklearn., Pandas fnⁿ

Functions → lambda, map, filter, reduce.

OOP → classes & obj, inheritance, Polymorphism.
method/fn, loop, break, cont
Reg ex etc.

Error Handling → try - except block.

Generators & Iterators.

Stats :

descriptive:

① Measure of Central Tendency → Mean, Median, Mode

② " " Spread → Variance, STD, IQR etc.

③ " " Symmetry → Skewness
" " Shape → Kurtosis

→ 5 pt. summary

- data type
- distribution
- No. of Categories (unique.)
- duplicates (df.duplicated().dn())
- Null value (df.isna(), sum())
- etc.

Inferential

- Null, Alt. Hypothesis
- P-value, & Significance Level)
- Z-score
- T-table, Z-table
- T-test, Z-test
- f-test (ANOVA)
- ANCOVA ,
- chi-Square
- CLT
- Empirical formula.

Statistics

Prob

- Bayes' Theorem
- Problem
- P & C.

Linear Algebra

- Matrix
- Vector
- Angle between vector

→ Eigen value & Eigen vectors.

(ML)

Regression:

Algo → Lin., DT, RF, ...

Assumption → " " "

Evaluation → R^2 , Adj- R^2

MSE, RMSE, MAE etc.

Hyperparameters → max_depth, max_features

Classification:

Algo → DT, RF, Logistic R, KNN, SVC etc..

Assumption → " "

Evaluation → Acc., Miscln., Prec. Rec.
AUC or ROC, AUC of PRC
F1-score etc.

diff betw TP, FP, TN, FN

Type I & Type II error

Hyperparameters → k

details of each algo

① RF? → Spltg with replace

② k in KNN? →

③ Kernel trick SVM? →

- ④ K in KNN Selection ?
- ⑤ K in K-means ?
how to select it ?
- ⑥ How Gradient descent works ?

1st Round

Python ✓

& Basic ML, SQL

2nd round

ML

& Projects

& Python. DB

3rd round

Magnitude → problem solving skills

4th round

Magnitudal

~~NLP & DL.~~