

Mock Interview Topics and Questions

(Additional questions to those already in LMS.)

Introduction.

Projects and data science understanding

explanation of projects completed

Business case studies

<https://towardsdatascience.com/the-ultimate-guide-to-cracking-business-case-interviews-for-data-scientists-part-1-cb768c37edf4>

<https://towardsdatascience.com/tagged/case-study>

<https://data-flair.training/blogs/data-science-case-studies/>

<https://bigdata-madesimple.com/6-of-my-favorite-case-studies-in-data-science/>

data science scenario based questions / general data science techniques

credit card fraud prediction model

what are steps involved in data science project? What is EDA? Missing values, outlier treatment, univariate bivariate analysis, checking correlation

What is feature engineering? Types?

What is overfitting and underfitting and how to remove it?

What is variance, bias, variance-bias tradeoff?

Communication Skills

Should be intermediate

Statistics

what is statistics?

What are 2 branches of statistics?

What is central limit theorem?

what is hypothesis testing?

What is confidence interval?

what is p value?

what is mode and median?

What is chi square test?

what is z test?

what is normal distribution?

What is population and sample?

inter quartile range?

standard deviation?

outlier?

one tail and 2 tail hypothesis testing?

Supervised ML

What is bagging and boosting?

What is precision and recall?

What is accuracy formula?

What is confusion matrix?

What is information gain?

What is hyperparameter tuning?

Unsupervised ML

What is unsupervised ML? Examples of some business problems?

What is elbow curve?

What is PCA?

What is kmeans clustering? What is kNN?

SQL

Group by, case when, joins

which clause will you use to put filters on aggregated data?

Q1)

5 columns : date, customerid, product_quantity, price_per_item, region. Which customer id bought maximum products. Tablename is customers.

Q2)

5 columns : date, customerid, product quantity, price_per_item, region. Get sales during Christmas period (21 dec 2020 - 27 dec 2020) by region.

Q3)

6 columns : date, customerid, product quantity, price_per_item, region, transaction_id. Get count of transactions by region where customers had greater than 10000 \$ sales and less than 10000 \$ sales

Python

What are some examples of data structures in Python?

What are some examples of data types in Python?

difference between lists and tuples?

is python case sensitive?

what is indexing in Python?

with what number, does indexing start in Python?

How do you write comment in Python?

How do you display the rows and columns in a pandas dataframe?

Which function to use to join 2 datasets vertically?

Prepare pandas well

Miscellaneous

Additional questions :

These are some of the common data science interview questions;

What do you mean by word Data Science?

What is Data Visualization?

How you can define Data cleaning as a critical part of process?

Differentiate between Data modelling and Database design?

Describe in brief the data Science Process flowchart?

Compare and contrast R and SAS?

What do you understand by letter 'R'?

What all things R environment includes?

What are the applied Machine Learning Process Steps?

Compare Multivariate, Univariate and Bivariate analysis?

Differentiate between Uniform and Skewed Distribution?

What do you understand by term Transformation in Data Acquisition?

What do you understand by term Normal Distribution?

What is Data Collection?

What is Linear Regression?

.What do you understand by term Threshold limit value?

Differentiate between Validation Set and Test set?

How can R and Hadoop be used together?

What do you understand by term RIMPALA?

What do you understand by Big data?

What do you mean by Recommender systems?

What is K-Nearest Neighbour?
Does very less data lead to best model?
What is Pattern Recognition?
What are the major steps in exploratory data analysis?
What is Genetic Programming?
Difference Between Classification and Regression?
How to use labelled and unlabelled?
How to deal with unbalanced data?
If you have a smaller dataset, how would you handle?
What is Cluster Sampling?
Define False Positive and False Negative
Differentiate between Inductive and Deductive Learning
What is a confusion matrix?
What is Bagging and Boosting?
What are categorical variables?
When to use ensemble learning?
What is the trade-off between accuracy and interpretability?
What is a ROC Curve and How to Interpret It?
Explain PCA?
Differences between supervised and unsupervised learning?
Explain univariate, bivariate, and multivariate analyses.

<http://net-informations.com/ds/iq/default.htm>