Name of the Company for which you have appeared for Interview	Position for which you have applied?	Please mention all rounds in interview process	Hiring Challenge involved?
Sociosquares	Data scientist	Face to face, code test, HR round	NLP, web scraping questions and research paper reading
Shaadi.com	Data Scientist	Face to face, HR round	Technical Round
Barclays	ML Engineer	Round 1: Telephonic Interview:30 minutes :Project related questions , PCA and LDA, SVM. What kind of data would be required from the client for credit card fraud detection	Yes,
Maersk	Machine Learning Engineer	1: Resume Shortlisting and candidate shortlisting by HR. 2.20 minutes telephonic interview. 3. Online aptitude test. 4.Offsite aptitude test. 5. F2F tech round. 6. Managerial round	Yes, but based on R so did not attempt. Questions based on SQL were given.
Nielsen	Data Science Executive	1. Written Test based on aptitude, statistics, basic coding, SQL, Case study. 2.Hangout Tech Round. 3. Hangout Managerial Round. 4. Telephonic HR round	No
Difference-ai, Pentation Analytics	Intern	1	No
Nielsen	Executive - Data Science	1.Written Test 2. Technical Interview 3.Technical Interview 4.HR 5. Pymetric Test	No

Pentation analytics	Machine Learning Engineer	2 rounds and challenge test	Dataset of clients and mutual funds ,they wanted business insights
UbiSoft	R&D Engineer	3	yes, classify all the images in a given video of littleman and pinkpanther to be either little man or pink panther.
Mahindra	Data Scientist	HR and Technical Round	For a given problem statement, what metric to use Precision or Recall. In my case, he asked for Sales Propensity scenario.
QuaQua Enterprises Private Limited	Data Scientist	Telephonic round, Technical Assignment round and F2F in Hyderabad	Yes. The problem statement required to predict ratings based on user reviews for electronic items. Data was not given and asked to be scraped using Google Places API and extract user reviews and ratings.

Kantar Analytics	Data Scientist (R&D)	First round was with the Partner of Analytics Practice who asked my predominantly on my work experience. Second round was a technical round with the R&D Lead of Analytics who asked me questions only on Deep Learning.	No hiring challenge as it is an R&D role (as mentioned by them)
kaam.work	Data Scientist	HackerRank coding test, Video Assessment and Founder Zoom Interview	Yes, there was an online assessment on SQL, Python, Machine Learning and Deep Learning on HackerRank

Mahindra	Data Scientist	HR & Technical	No
Icruxsystem	Data scientist	Techinical round	No
ICRUXSYSTEM	Data Scientist	Technical	About Dataset which I used in my project
Digitas	Data Engineer and Data Scientist	1. Aptitude test 2. Technical Test 3. Technical interview 4. HR round	No
JAVIS	NLP Engineer	Assignment + Technical Telephonic + F2F Discussion with CEO	Yes, take-home was given which has been attached here.
GoAir	Data Analyst	Managerial round + Assignment	NA
Upstox	Data Analyst	Online Assessment Test on SQL ,Face to Face technical , HR Round	Yes

Episource	Data Science Associate	Technical telephonic+Assignment	NA
Transorg Analytics	Data Scientist	Technical Telephonic + Skype technical discussion + Business Telephonic + Final Discussion	NA

Edelweiss Insurance	Aws ML Engineer	Technical Assessment + Technical F2F	NA
Javis, Springer nature, Episource, LnT	Data Scientist	Technical Telephonic + Assignment + Technical F2F + Final Discussion	yes
KoiReader	NLP Engineer	Online Coding Round	NA
Accenture	Ml	Ml	Ml
Course5i, directi	Data analysis, data engineer	Aptitude test, HR, technical interview	NA
Chalhoub Group	Business Analyst	echnical Telephonic + Assignment + Technical F2F + Final Discussion	na

kaam.work	Data Scientist	HackerRank coding test, Video Assessment and Founder Zoom Interview	Yes, there was an online assessment on SQL, Python, Machine Learning and Deep Learning on HackerRank
Fractal	Data Scientist		
stockroom.io	Data scientist		
DataTeam			
Syndrome Technologies	Data Analyst		

All companies so far	Data Scientist	

ExZeo Noida, Dainik Bhaskar, Profisor Services, Flexiloans	Data Scientist		
Episource	Data Scientist		
Fork Media	Data Scientist		

Differnet companies	Data Scientist		
Episource,ForkMedia,Fractal Analytics	Data Scientist/Analyst		

Online Aptitude test?	Online Coding test?
No	No
No	No
No	No
Yes, basic logical reasoning questions	No
No	No
No	No
There was a written pen and paper Aptitude Test. It involved True and False questions, Short answer questions , Coding Questions , Plotting graphs question, Math questions and Case studies. Largely it was blend of Stats and coding language proficiency test.	No

No	No
yes, online test was on python	yes, 1. In a given string find out all the consecutive letters, for example if letter "hello" is given output should be {'l':2} if string is "hellllo" --> {'l':4} if words are not consecutive then those words won't be counted. if more than one characters are repeating then those should also also be printed. if one character is consecutive more than once then maximum value should be printed. for ex if "rrorrr" --> {'r':3} 2. Give a list of numbers for a histogram, find out max rectangle in that histogram.
No	No
No	No coding test, the assignment was there for checking the coding aspect

None	None
Yes, there was an online assessment on HackerRank which contained 75 MCQ questions to be solved within 90 minutes. The test covered skills on SQL, Python (for data science), linear algebra and calculus, Machine Learning (scikit-learn, pandas), data modelling, data cleaning, feature engineering and Deep Learning (TensorFlow 1.0). Question difficulty level ranged from easy to medium (5-10 were hard).	Same as above answer

No	No
No	No
No	No
There was an offline aptitude test for 40 minutes with 10 questions. They were more of general aptitude especially to test linear equations.	No
NA	NA
NA	NA
NA	YES

NA	yes
NA	NA

NA	Yes
No	No
NA	Yes
Yes	Yes
Yes	NA

Yes, there was an online assessment on HackerRank which contained 75 MCQ questions to be solved within 90 minutes. The test covered skills on SQL, Python (for data science), linear algebra and calculus, Machine Learning (scikit-learn, pandas), data modelling, data cleaning, feature engineering and Deep Learning (TensorFlow 1.0). Question difficulty level ranged from easy to medium (5-10 were hard).	Same as above answer

Questions that were asked in Technical Rounds?
Difference between linear regression and logistic regression. What are the good evaluation methods..about the project mention on resume
What is the drawback of PCA, why dimension Reduction is important,Difference application of Regression, Statistics questions, what is Regularization, why over fitting happen.
--
Why Data Science? Linear,Logistic Regression,Decision Trees. Regularization Methods. In Depth discussion on AUC ROC curve. K-Means. Why switching to R from python?
1. How do you rotate a survey of questionnaire in a locality of 50 buildings and 20 houses each? 2. Linear regression , Regularization Methods. 3. Various Sampling Methods 4.Types of distributions 5.Basic Descriptive stats questions 6. Why switching to a Market Research firm from a technical background?
Na
It was largely based on my CV as I had two internships in Market Research field. Apart from that : - Basic concepts of Stats like what is descriptive and inferential statistics ,Correlation, Distributions , regression ,Sampling techniques etc. - Basic coding operation using python like merge,pivot etc

Lasso regression , logistic regression
Linear Regression, Decision Tree, Linear regression example, time series example, whether weather forecasting a linear regression problem or not.
1. Explaination of the work done so far 2. Approaches used in Hackathon 3. Questions on SVM 4. Questions on Precision, Recall and F1 score 5. Multicollinearity - where to use and why to use
There were two case studies. The first question was asked "How would you go about designing a personalised travel itinerary?" Second question was "Which online subscription service (Netflix, Hotstar, Amazon Prime) would you improve technically and how?" It also involved presenting the assignment and asked about data collection challenges and my approach to the problem statement.

I can only detail the questions asked in the technical telephonic round by the R&D Lead of Analytics. They are:
1. Describe the most challenging project you have worked on.
2. Why will you chose ResNet or Inception architecture over AlexNet? ---> ANSWER: Because of vanishing gradient and skip connections
3. What is vanishing gradients and which activation functions have it?
4. What is the problem with ReLU as activation functions and how will you overcome it? ----> ANSWER: ReLU suffers from dying gradients when inputs are negative. One possible solution is using sophisticated activation functions like Leaky ReLU.
5. Why ReLU is so frequently used in CNN architecture? ----> ANSWER: Because ReLU activation implies certain features (like nose, ears, eyes etc.) are being detected by specific neurons.
6. Suppose you have a simple 3 layer neural network with no activation functions. How will this network behave and why so? ----> ANSWER: This network will behave as a linear model because the output layer is a linear function of the hidden layer and the input layer, the hidden layer is a linear function of the inputs.
7. Lets assume you have a linear model and your R-squared is 0.76. Now you add a random feature to it. What will happen to the value of R-squared? Is it a good practice to do? What will you do if it is not a good practice? ----> ANSWER: The R-squared value will either remain the same or will increase if a new feature is added. It is not a good practice to add a random feature just to boost the evaluation metric. I will most likely look for past studies on the same subject and examine them for relevant features to improve my evaluation metric. Also, I

The Founder interview consisted of three questions:
1. Which of the following phenomena is/are undesirable in linear regression?
(a) Auto-correlation (b) Multicollinearity (c) Target value is continuous
----> ANSWER: (a) and (b); Assumptions of linear regression
2. Given a sample of 100 students whose average is 70 and standard deviation is 16, what will happen to average and standard deviation if the sample size is reduced to half i.e. 50? ----> ANSWER: Average will be half i.e. 35 and standard deviation will be square root of 16 i.e. 4
3. Which of the following techniques cannot be used for outlier detection?
(a) K-Means (b) SVM (c) PCA (d) All of the above
----> ANSWER: (c) PCA

Questions were also asked on my previous projects.

The discussion was mainly about the projects that were mentioned in my resume. I was asked about my approach to solving those projects. I have mentioned Ad clicks classification project that we did in one of the hackathons. The questions were about the algorithm that I used for that problem and why we selected that algorithm. They also asked about the metric that we used for evaluation. There were a few general questions regarding precision, recall and accuracy.
Python basic programing question and questions about project mentioned in resume
Validation metric used
About the projects that I have worked on.
Suppose you work for Twitter and want to create a system that would give suggestions for hashtag after writing the tweet. How would you go about creating this system ? Explain right from the data collection part. Explain any ML model. Explain bias-variance trade-off. Is logistic regression a linear model ? Explain lasso and ridge, why are they used and their difference. What is the cost function for logistic regression ? What is p-value and significance value - how to decide which hypothesis to go for ? What is pruning in decision trees? What parameters decide pruning in decision trees ? What dimensionality reduction technique to go for ? Explain PCA. What is the formula for cosine similarity.
Python Assessment on the given data & questions on profile description as per resume.
What is Data Mining, What is Data Profiling, Steps involved in Projects you have done in Past.

Describe the projects undertaken. For a specific project chosen by the interviewer, I was asked to describe the procedure from start to end. Whatever technical terms were used by me in this description, I was grilled on that. The following is a list of questions I was asked based on my project: 1. Is the decision boundary of logistic regression linear? 2. What are the assumptions for logistic regression and how would you check for each one 3. What are the different sampling techniques and difference between each 4. What is the drawback of SMOTE 5. What was the basis for using oversampling instead of undersampling 6. Description of different feature selection methods 7. Is backward selection a recursive feature elimination technique? 8. Details of ANOVA 9. What are generators and yields in python? 10. Different ways of defining a function and when to use which 11. What is Apply, map, filter and when to use which 12. Have you used classes in the project
Explain about the projects? What Feature engineering you did? Explain any one algorithm in detail? What is difference between bagging & boosting? How do you select a best model ? What Model validation metric you have used? Explain the metrics? How did you come up with target variable in classification project?(specific to business domain) How do you check feature importance? Categorical variables treatment? Data imbalance handling? What is bias variance tradeoff? Explain the same in bagging n boosting? How to save a trained model ? How to deploy? Assumptions of linear regression?

Questions related to project Scenario based questions Cost function of logistic regression Interpretation of coefficients Bias variance tradeoff Multicolinearity tackling Assumptions of linear regression How to calculate R squared in logistic regression Explain k means clustering Write program of k means from scratch(30 min time) How to tackle with assumptions of linear regression Bagging and boosting based questions and math behind it Model deployment Aws Text data based questions Sql questions Types of joins, CTE , subqueries, diff between in and exists, union, union all
- What is a normal distribution? - What is the p-value? - What is a central limit theorem? - What is naive in naive Bayes? - Explain the mean, median, mode. - What are precision and recall? - What is the difference between bagging and boosting? - Different between MSE and MAE? When to use which matrics? - Types of linear regression - Does logistic regression use regularization? - Can we use a decision tree for regression? - How to choose the best features from 1000 features without using PCA? - suppose you have given logistic regression, SVM, random forest and xgboost and all of them having accuracy between 88 - 90 %. For model deployment which model you are going to deploy. - What is POS tagging? - Detailed flow of the tweeter sentiment analysis. Which algo you are going to use for this logistic or SVM? - Which algorithm you are comfortable with and Explain that in detail. - suppose you have given the 1000 resumes in order to select the best candidate among them. Which way you are going to solve the problem. - Three Assumptions of Linear regression?
NA (as I could only appear in coding round, the question sheet I have attached)
F2f
SQL and python based
You can club telephonic and F2F technical round questions and share all

The Founder interview consisted of three questions: 1. Which of the following phenomena is/are undesirable in linear regression? (a) Auto-correlation (b) Multicollinearity (c) Target value is continuous ----> ANSWER: (a) and (b); Assumptions of linear regression 2. Given a sample of 100 students whose average is 70 and standard deviation is 16, what will happen to average and standard deviation if the sample size is reduced to half i.e. 50? ----> ANSWER: Average will be half i.e. 35 and standard deviation will be square root of 16 i.e. 4 3. Which of the following techniques cannot be used for outlier detection? (a) K-Means (b) SVM (c) PCA (d) All of the above ----> ANSWER: (c) PCA Questions were also asked on my previous projects.
Machine Learning algorithms you know?
Explain Haptik Project you worked on?
How does Cost Function work?
Difference between count vectorizer and tfidf?
Bias variance trade-off
L1 and L2 regularization
Question Asked on Haptik project
Difference between K-NN and K-means
Case study - If a guy collects all the newspaper of all publications in the 2016 year then if 2mrw some other guy comes with a newspaper that does not have date and publication on that then how will you identify when was that newspaper published in 2016 year.
Difference between Supervised and Unsupervised
questions on projects
case study - a two wheeler company needs recommendation for setting up 10 new dealerships for a new vehicle, we had details of the previous dealerships at different places (cities, towns) all over the country, this new dealership sales should be 15% more than the previous year sales, how do you rank the dealership if you have to rank them in an order, how will you validate your recommendations?
questions on the approach and reasoning behind at differents steps for the test submitted
sigma levels range and accuracy
what is a t distribution is it related to t test
how do you detect the outliers?
Which of the following are sensitive to outliers: Mean or Median? Why?
How is a split made in a Decision Tree?

When do you use Gini Index or Entropy?
Does Decision Tree have high bias or variance?
What is Bias - Variance Tradeoff
Why is Random Forest so special?
DIFFICULT - When the basis of splitting data in Decision Tree is information gain and for the case of pruning we again prune the full grown tree based on information gain. Why invest so much time in 1st growing the tree based on info gain and then again pruning the tree based on the same parameter?
How many features are selected when creating decision trees in a Random forest?
Derive stochastic gradient descent in python?
Probability questions on Baye's Theorem
What is p-value and confidence interval? DIFFICULT - P-value meaning in Linear Regression?
DIFFICULT - What is PCA? Difference between PCA & SVD? Can SVD be used for Dimensionality Reduction instead of PCA? Difference between PCA & Factor Analysis?
DIFFICULT - Given a 8x8 Matrix consiting of 0's and 1's representing the shape of Number '8'. Apply PCA on it and describe the nature of the output. How many components will you chose
L1 and L2 regularization in depth
TRICKY - Why do you need Logistic Regression if we have linear regression? Drawbacks of using Linear Regression for classification
What is eigen value and eigen vector? How are they useful in PCA?
When do we use SVD over PCA?
TRICKY - How do you tackle logistic regression with low probability events? Ans - Penalized Likelihood algorithm
In linear regression model, the order is 1, how do you say that the model is overfitting when we explain that the coefficients of higher order polynomial terms are huge in overfitting?
Finding out the distance between support vectors and the margin in 4-D plane
TRICKY - K-means algorithm is for numeric data clustering. How do we tackle clustering if we have categorical data? Ans -> K-mode algorthim helps in that case.
Preprocessing in NLP, Some questions on creating Regex patterns. Difference between tf-idf and count vectorizer.
Write code for n-gram implementation in Python

TRICKY - Why is numpy and pandas used in Python? Why can't we use Base python for Data Science? The word that interviewer wanted to hear was - Pandas and Numpy introduce parallelism with matrix operations
Why is ROC Curve used? Interpreting the ROC-AUC Curve
what is CART?
Difference between Linear and logistic Regression
What is Bias - Variance Tradeoff
how to avoid overfitting
how is random forest diffrerent from decision tree
why pruning is done
how does gradient descent work and why it is needed?
Tell me something about yourself
Tell me about your work experience in depth
Tell me about the project you have done in your job
Explain how would you approach a problem given to you
Why is logistic regression called regression even though it's classification
What do you know about decision trees
What is difference between random forest and gradient boost
Have you ever deployed a model in production
After a model has been developed how do you use in a flask or front end environment
Have you ever worked on hadoop? Can you explain about it.
Tell me about the projects you have worked on
Explain your approach to the problem
Explain confusion matrix
Do you know mongo db
How do you label your data
How do you measure performance of your model once it is in production
How do you label your data
How do you measure performance of your model once it is in production
Tell me about projects you have worked on and your experience in detail.
Explain linear regression ?
Functions of pandas and numpy?
What is scikit learn and contents of scikit learn why it is used ?
Combining two datasets using pandas syntax.
Consider a cancer detection model if it has the accuracy of 90% then what will be the rmse and is the 90% accuracy is good or bad ?

Difference between normal python list structures and numpy.
Difference between different Python IDE's used for data science and analysis which you are aware of.
Queries of SQL.
Can we perform linear regression using decision tree?
Difference between loss function and evaluation metrics ?
Calculate the gini index or entropy and find the next node.
Explain the summary report output of logistic regression
Difference between decision tree and random forest?
How does the random forest selects node ?
How to tranform the grouped row values to columns as a feature in SQL ?
Fractal test problems from hackerRank.
What is descriptive and inferential staistics explain shortly.
What techniques to handle the imbalance data ?
How to handle the large size data using pandas consider the data size around 10GB's or more.
Explain confusion matrix,precision,recall.
Concepts of Linear Regression- Homoscedascity,Variable Inflation Factor,P-value
Feature Engineering concepts - how to handle ,outliers,missing values
Difference beween random forest and gradient boosting
How is RNN different from a traditional neural network
Why is logistic regression called regression when it's a classification technique
Assumptions of linear regression
Why ROC doesnt work incase of imbalanced datasets
How does a Decision tree work - ID3 Algorithm
Case studies - Sales forecasting in the retail industry,churn predictions
how to handle large datasets
Basically they ask about the projects in detail - where they mainly focus on how you did feature engineering and feature selection