Data Science

Introduction to Data Science

Learn how to prepare for the wide range of questions that come up in data science interviews.

0 of 8 Completed

Easy SQL Questions

Get started on tackling easy level SQL questions involving aggregations, joining multiple tables, and pulling data for beginning analytical reports.

0 of 12 Completed

Medium SQL Questions

Medium level SQL questions utilize more advanced concepts like sub-queries, window functions, and solving case study problems.

0 of 19 Completed

Hard SQL Questions

Let's tackle advanced SQL interview questions that focus on multi-joins and layers of data interpretation. These questions may come up in take-home challenges and senior level interviews.

0 of 9 Completed

Data Structures

Data structures in Python attempt to be more intuitive and flexible than traditional data structures in other programming languages.

0 of 9 Completed

Common DS Packages

As we said in the first section of this course, a major benefit of using Python for data science in comparison to other programming languages is the availability of a large number of useful packages that are distributed under a free license.

0 of 11 Completed

Python Questions: Hard

Let's try some hard Python questions that you would see in tougher data science interviews and many machine learning interviews.

0 of 6 Completed

Basic Probability

Probability Theory is the branch of mathematics that deals with uncertainty, underpinning all of statistics and machine learning.

0 of 10 Completed

Discrete Distributions

All areas of study in math can roughly be divided into two camps: discrete mathematics and continuous mathematics. Perhaps the best way to describe the difference between the two is to talk about what each of the branches means by "number."

0 of 12 Completed

Continuous Distributions

Continuous probability distribution: A probability distribution in which the random variable X can take on any value (is continuous).

0 of 6 Completed

Sampling Theorems

Thus far in this course, we have considered random variables under an idealized scenario where we know the distribution of the random variable.

0 of 7 Completed

Hypothesis Testing

Hypothesis testing covers the fundamental theory and background behind A/B Testing. In this course we'll cover Z and T test, multiple hypothesis testing, and the different type errors.

0 of 11 Completed

Confidence Intervals

Confidence intervals help us deal with this imprecision by giving us a way to talk about a range of values with some certainty where the true value of the statistic is contained in.

0 of 6 Completed

A/B Testing & Experiment Design

Let's start with a general framework for A/B testing. In practice, an A/B testing and experimentation all follow a step by step process of setting metrics and designing experiments.

0 of 10 Completed

A/B Testing Common Scenarios

The next couple of chapters will cover common scenarios and concepts involved in A/B testing. As A/B testing involves statistical concepts, there may be terms that you need refreshing on.

0 of 9 Completed

A/B Testing Tradeoffs

There are scenarios where A/B testing is not necessarily the best course of action. Often, there are technical, infrastructure, or practical concerns that come up while planning an A/B test.

0 of 6 Completed

Statistics

This is a refresher on some important statistical concepts that will help us with A/B testing and beyond. While by no means a comprehensive guide, this chapter will go over some important basics about statistical testing and probability distributions.

0 of 11 Completed

Data Analytics Fundamentals: Causal Inference

In this course we'll go over the core concepts of causality, significance, and analyzing data. This is meant as a quick refresher and a high level overview of causal inference basics to eventually apply them in data analytics problems.

0 of 9 Completed

Diagnosing and Investigating Metrics

Investigating metrics is a type of product intuition problem that will come up frequently in interviews. Examples of this are typically phrased along the lines of - If X metric is up/down by Y percent, how would you investigate it?

0 of 12 Completed

Measuring Success

Measuring the success of products is critical to data science and analytics interviews. Generally, this question is an encapsulation of every time a product manager or executive asks the question: “So, how is it doing?”.

0 of 11 Completed

Feature Change

Before launching a feature, we can imagine that the first step we’d have to take is analyzing the existing data in our product to make a decision about exactly what to build. This process is what creates the building or change of a feature problem that gets asked in product interviews.

0 of 10 Completed

Metric Trade-Offs

Metric trade-off type questions can occur on their own in product interviews or as part of a larger product or AB testing interview discussion.

0 of 4 Completed

Modeling Case Study

The machine learning and modeling case study is the most common type of interview question that tests a combination of modeling intuition and business application.

0 of 2 Completed

Data Pre-Processing

Data processing and analysis is the first step that we need to consider once we've clarified details and started down the path of building the model.

0 of 5 Completed

Feature Selection

Feature selection and feature engineering is the second part of the data processing step. Once we've understood what our data looks like, we need to begin to theorize the kinds of features we would use to build the model.

0 of 4 Completed

Model Selection

Model selection is usually the crux of any modeling case study problem. We want to be able to select a model or machine learning algorithm that will combine a bunch of factors to become the most optimal algorithm for the problem.

0 of 4 Completed

Machine Learning Algorithms

We have touched on the different machine learning algorithms throughout this lesson, but haven't yet dived deep into each one. The prior for this course is that you, as a candidate, have an idea of basic machine learning concepts, and the different modeling algorithms are one such example of them.

0 of 7 Completed

Model Evaluation

Most machine learning model deployment requires some technical details and implementation to doing so. But we can abstract away from that in an interview when we're focusing on the model roll out.

0 of 9 Completed

Applied Modeling

Applied modeling is a type of case question asked about practical machine learning. The most common type of question framework is: Given an example scenario with a machine learning system or model, how would you analyze and fix the problem?

0 of 5 Completed

Generalized Linear Models and Regression

Regression models are used to predict the value of a dependent variable from one or more independent variables.

Courses

Courses in this learning path are:

Introduction to Data Science

Introduction to Data Science

Learn how to prepare for the wide range of questions that come up in data science interviews.

0 of 8 Completed

Easy SQL Questions

Easy SQL Questions

Get started on tackling easy level SQL questions involving aggregations, joining multiple tables, and pulling data for beginning analytical reports.

0 of 12 Completed

Medium SQL Questions

Medium SQL Questions

Medium level SQL questions utilize more advanced concepts like sub-queries, window functions, and solving case study problems.

0 of 19 Completed

Hard SQL Questions

Hard SQL Questions

Let's tackle advanced SQL interview questions that focus on multi-joins and layers of data interpretation. These questions may come up in take-home challenges and senior level interviews.

0 of 9 Completed

Data Structures

Data Structures

Data structures in Python attempt to be more intuitive and flexible than traditional data structures in other programming languages.

0 of 9 Completed

Common DS Packages

Common DS Packages

As we said in the first section of this course, a major benefit of using Python for data science in comparison to other programming languages is the availability of a large number of useful packages that are distributed under a free license.

0 of 11 Completed

Python Questions: Hard

Python Questions: Hard

Let's try some hard Python questions that you would see in tougher data science interviews and many machine learning interviews.

0 of 6 Completed

Basic Probability

Basic Probability

Probability Theory is the branch of mathematics that deals with uncertainty, underpinning all of statistics and machine learning.

0 of 10 Completed

Discrete Distributions

Discrete Distributions

All areas of study in math can roughly be divided into two camps: discrete mathematics and continuous mathematics. Perhaps the best way to describe the difference between the two is to talk about what each of the branches means by "number."

0 of 12 Completed

Continuous Distributions

Continuous Distributions

Continuous probability distribution: A probability distribution in which the random variable X can take on any value (is continuous).

0 of 6 Completed

Sampling Theorems

Sampling Theorems

Thus far in this course, we have considered random variables under an idealized scenario where we know the distribution of the random variable.

0 of 7 Completed

Hypothesis Testing

Hypothesis Testing

Hypothesis testing covers the fundamental theory and background behind A/B Testing. In this course we'll cover Z and T test, multiple hypothesis testing, and the different type errors.

0 of 11 Completed

Confidence Intervals

Confidence Intervals

Confidence intervals help us deal with this imprecision by giving us a way to talk about a range of values with some certainty where the true value of the statistic is contained in.

0 of 6 Completed

A/B Testing & Experiment Design

A/B Testing & Experiment Design

Let's start with a general framework for A/B testing. In practice, an A/B testing and experimentation all follow a step by step process of setting metrics and designing experiments.

0 of 10 Completed

A/B Testing Common Scenarios

A/B Testing Common Scenarios

The next couple of chapters will cover common scenarios and concepts involved in A/B testing. As A/B testing involves statistical concepts, there may be terms that you need refreshing on.

0 of 9 Completed

A/B Testing Tradeoffs

A/B Testing Tradeoffs

There are scenarios where A/B testing is not necessarily the best course of action. Often, there are technical, infrastructure, or practical concerns that come up while planning an A/B test.

0 of 6 Completed

Statistics

Statistics

This is a refresher on some important statistical concepts that will help us with A/B testing and beyond. While by no means a comprehensive guide, this chapter will go over some important basics about statistical testing and probability distributions.

0 of 11 Completed

Data Analytics Fundamentals: Causal Inference

Data Analytics Fundamentals: Causal Inference

In this course we'll go over the core concepts of causality, significance, and analyzing data. This is meant as a quick refresher and a high level overview of causal inference basics to eventually apply them in data analytics problems.

0 of 9 Completed

Diagnosing and Investigating Metrics

Diagnosing and Investigating Metrics

Investigating metrics is a type of product intuition problem that will come up frequently in interviews. Examples of this are typically phrased along the lines of - If X metric is up/down by Y percent, how would you investigate it?

0 of 12 Completed

Measuring Success

Measuring Success

Measuring the success of products is critical to data science and analytics interviews. Generally, this question is an encapsulation of every time a product manager or executive asks the question: “So, how is it doing?”.

0 of 11 Completed

Feature Change

Feature Change

Before launching a feature, we can imagine that the first step we’d have to take is analyzing the existing data in our product to make a decision about exactly what to build. This process is what creates the building or change of a feature problem that gets asked in product interviews.

0 of 10 Completed

Metric Trade-Offs

Metric Trade-Offs

Metric trade-off type questions can occur on their own in product interviews or as part of a larger product or AB testing interview discussion.

0 of 4 Completed

Modeling Case Study

Modeling Case Study

The machine learning and modeling case study is the most common type of interview question that tests a combination of modeling intuition and business application.

0 of 2 Completed

Data Pre-Processing

Data Pre-Processing

Data processing and analysis is the first step that we need to consider once we've clarified details and started down the path of building the model.

0 of 5 Completed

Feature Selection

Feature Selection

Feature selection and feature engineering is the second part of the data processing step. Once we've understood what our data looks like, we need to begin to theorize the kinds of features we would use to build the model.

0 of 4 Completed

Model Selection

Model Selection

Model selection is usually the crux of any modeling case study problem. We want to be able to select a model or machine learning algorithm that will combine a bunch of factors to become the most optimal algorithm for the problem.

0 of 4 Completed

Machine Learning Algorithms

Machine Learning Algorithms

We have touched on the different machine learning algorithms throughout this lesson, but haven't yet dived deep into each one. The prior for this course is that you, as a candidate, have an idea of basic machine learning concepts, and the different modeling algorithms are one such example of them.

0 of 7 Completed

Model Evaluation

Model Evaluation

Most machine learning model deployment requires some technical details and implementation to doing so. But we can abstract away from that in an interview when we're focusing on the model roll out.

0 of 9 Completed

Applied Modeling

Applied Modeling

Applied modeling is a type of case question asked about practical machine learning. The most common type of question framework is: Given an example scenario with a machine learning system or model, how would you analyze and fix the problem?

0 of 5 Completed

Generalized Linear Models and Regression

Generalized Linear Models and Regression

Regression models are used to predict the value of a dependent variable from one or more independent variables.

DS Applied

We can't lie - Data Science Interviews are TOUGH. Especially tricky - probability and statistics questions asked by top tech companies & hedge funds during the Data Science, Data Analyst, and the Quant Trading Interview process.

That's why we put together 40 real probability & statistics data science interview questions asked by companies like Facebook, Amazon, Two Sigma, & Bloomberg. We have solutions to all 40 problems, and to 161 other data interview problems on SQL, Machine Learning, and Product/Business Sense in our book, Ace The Data Science Interview. You can also practice some of these same exact questions on DataLemur's statistics interview questions section.

DataLemur has hundreds of real Statistics and Probability Interview questions, sourced from real Data Science and Data Analyst interviews at companies like Facebook and Google.

So, without further ado, here are:

the probability & stat concepts to review before your DS interview

20 probability questions asked by top tech-companies & Wall Street

20 statistics questions asked by FANG & Hedge Funds

solutions to 5 of the probability questions

solutions to 5 of the statistics questions

links to more data science interview resources

Probability & Statistics Concepts To Review Before Your Data Science Interview

Because probability & statistics is foundational to the field of Data Science, before the interview you should review:

Probability Basics & Random Variables

Probability Distributions

Hypothesis Testing

Regression Analysis

In case these statistical concepts sound alien to you, check out some of our favorite Statistics Books for Data Analysts to get a gentle refresher.

Probability Basics and Random Variables

The beginnings of probability start with thinking about sample spaces, basic counting and combinatorial principles. Although it is not necessary to know all of the ins-and-outs of combinatorics, it is helpful to understand the basics for simplifying problems. One classic example here is the “stars and bars” counting method.

The other core topic to study is random variables. Knowing concepts related to expectation, variance, covariance, along with the basic probability distributions is crucial.

Probability Distributions

For modeling random variables, knowing the basics of various probability distributions is essential. Understanding both discrete and continuous examples, combined with expectations and variances, is crucial. The most common distributions discussed in interviews are the Uniform and Normal but there are plenty of other well-known distributions for particular use cases (Poisson, Binomial, Geometric).

Most of the time knowing the basics and their applications should suffice. For example, which distribution would flipping a coin be under? What about waiting for an event? It never hurts being able to do the derivations for expectation, variance, or other higher moments.

Hypothesis Testing

Hypothesis testing is the backbone behind statistical inference and can be broken down into a couple of topics. The first is the Central Limit Theorem, which plays an important role in studying large samples of data. Other core elements of hypothesis testing: sampling distributions, p-values, confidence intervals, type I and II errors. Lastly, it is worth looking at various tests involving proportions, and other hypothesis tests.

Most of these concepts play a crucial role in A/B testing, which is a commonly asked topic during interviews at consumer-tech companies like Facebook, Amazon, and Uber. It's useful to not only understand the technical details but also conceptually how A/B testing operates, what the assumptions are, possible pitfalls, and applications to real-life products.

Modeling

Modeling relies on a strong understanding of probability distributions and hypothesis testing. Since it is a broad term, we will refer to modeling as the areas which have a strong statistical intersection with Machine Learning. This includes topics such as: linear regression, maximum

likelihood estimation, & bayesian statistics. For interviews focused on modeling and machine learning, knowing these topics is essential.

20 Probability Interview Problems Asked By Top-Tech Companies & Wall Street

[Facebook - Easy] [Coin Fairness Test on DataLemur] There is a fair coin (one side heads, one side tails) and an unfair coin (both sides tails). You pick one at random, flip it 5 times, and observe that it comes up as tails all five times. What is the chance that you are flipping the unfair coin?

[Lyft - Easy] You and your friend are playing a game. The two of you will continue to toss a coin until the sequence HH or TH shows up. If HH shows up first, you win. If TH shows up first, your friend wins. What is the probability of you winning?

[Google - Easy] What is the probability that a seven-game series goes to 7 games?

[Facebook - Easy] Facebook has a content team that labels pieces of content on the platform as spam or not spam. 90% of them are diligent raters and will label 20% of the content as spam and 80% as non-spam. The remaining 10% are non-diligent raters and will label 0% of the content as spam and 100% as non-spam. Assume the pieces of content are labeled independently from one another, for every rater. Given that a rater has labeled 4 pieces of content as good, what is the probability that they are a diligent rater?

[Bloomberg - Easy] Say you draw a circle and choose two chords at random. What is the probability that those chords will intersect?

[Amazon - Easy] 1/1000 people have a particular disease, and there is a test that is 98% correct if you have the disease. If you don't have the disease, there is a 1% error rate. If someone tests positive, what are the odds they have the disease?

[Facebook - Easy] There are 50 cards of 5 different colors. Each color has cards numbered between 1 to 10. You pick 2 cards at random. What is the probability that they are not of same color and also not of same number?

[Tesla - Easy] A fair six-sided die is rolled twice. What is the probability of getting 1 on the first roll and not getting 6 on the second roll?

[Facebook - Easy] What is the expected number of rolls needed to see all 6 sides of a fair die?

[Microsoft - Easy] Three friends in Seattle each told you it's rainy, and each person has a 1/3 probability of lying. What is the probability that Seattle is rainy? Assume the probability of rain on any given day in Seattle is 0.25.

[Uber - Easy] Say you roll three dice, one by one. What is the probability that you obtain 3 numbers in a strictly increasing order?

[Bloomberg - Medium] Three ants are sitting at the corners of an equilateral triangle. Each ant randomly picks a direction and starts moving along the edge of the triangle. What is the probability that none of the ants collide? Now, what if it is k ants on all k corners of an equilateral polygon?

[Two Sigma - Medium] What is the expected number of coin flips needed to get two consecutive heads?

[Amazon - Medium] How many cards would you expect to draw from a standard deck before seeing the first ace?

[Robinhood - Medium] A and B are playing a game where A has n+1 coins, B has n coins, and they each flip all of their coins. What is the probability that A will have more heads than B?

[Airbnb - Medium] Say you are given an unfair coin, with an unknown bias towards heads or tails. How can you generate fair odds using this coin?

[Quora - Medium] Say you have N i.i.d. draws of a normal distribution with parameters μ and σ. What is the probability that k of those draws are larger than some value Y?

[Spotify - Hard] A fair die is rolled n times. What is the probability that the largest number rolled is r, for each r in 1..6?

[Snapchat - Hard] There are two groups of n users, A and B, and each user in A is friends with those in B and vice versa. Each user in A will randomly choose a user in B as their best friend and each user in B will randomly choose a user in A as their best friend. If two people have chosen each other, they are mutual best friends. What is the probability that there will be no mutual best friendships?

[Tesla - Hard] Suppose there is a new vehicle launch upcoming. Initial data suggests that any given day there is either a malfunction with some part of the vehicle or possibility of a crash, with probability p which then requires a replacement. Additionally, each vehicle that has been around for n days must be replaced. What is the long-term frequency of vehicle replacements?

20 Statistics Problems Asked By FAANG & Hedge Funds

[Facebook - Easy] How would you explain a confidence interval to a non-technical audience?

[Two Sigma - Easy] Say you are running a multiple linear regression and believe there are several predictors that are correlated. How will the results of the regression be affected if they are indeed correlated? How would you deal with this problem?

[Uber - Easy] Describe p-values in layman's terms.

[Facebook - Easy] How would you build and test a metric to compare two user's ranked lists of movie/tv show preferences?

[Microsoft - Easy] Explain the statistical background behind power.

[Twitter - Easy] Describe A/B testing. What are some common pitfalls?

[Google - Medium] How would you derive a confidence interval from a series of coin tosses?

[Stripe - Medium] Say you model the lifetime for a set of customers using an exponential distribution with parameter λ, and you have the lifetime history (in months) of n customers. What is your best guess for λ?

[Lyft - Medium] Derive the mean and variance of the uniform distribution U(a, b).

[Google - Medium] Say we have X ~ Uniform(0, 1) and Y ~ Uniform(0, 1). What is the expected value of the minimum of X and Y?

[Spotify - Medium] You sample from a uniform distribution [0, d] n times. What is your best estimate of d?

[Quora - Medium] You are drawing from a normally distributed random variable X ~ N(0, 1) once a day. What is the approximate expected number of days until you get a value of more than 2?

[Facebook - Medium] Derive the expectation for a geometric distributed random variable.

[Google - Medium] A coin was flipped 1000 times, and 550 times it showed up heads. Do you think the coin is biased? Why or why not?

[Robinhood - Medium] Say you have n integers 1...n and take a random permutation. For any integers i, j let a swap be defined as when the integer i is in the jth position, and vice versa. What is the expected value of the total number of swaps?

[Uber - Hard] What is the difference between MLE and MAP? Describe it mathematically.

[Google - Hard] Say you have two subsets of a dataset for which you know their means and standard deviations. How do you calculate the blended mean and standard deviation of the total dataset? Can you extend it to K subsets?

[Lyft - Hard] How do you randomly sample a point uniformly from a circle with radius 1?

[Two Sigma - Hard] Say you continually sample from some i.i.d. uniformly distributed (0, 1) random variables until the sum of the variables exceeds 1. How many times do you expect to sample?

[Uber - Hard] Given a random Bernoulli trial generator, how do you return a value sampled from a normal distribution

---

The role of a Meta Data Scientist Product analytics

Data scientist, Product analytics is a vital role at Meta. It focuses heavily on business problems. As a DS, Product Analytics you will work to bring out the best product and market analysis to help Meta make data-driven business decisions. For related roles at similarly levelled companies, check out the Lyft Data Scientist and Microsoft Data Scientist guides.

The Data Scientist Product Analytics role has work across the following four areas:

Product Operations: Forecasting and setting product team goals. Designing and evaluating experiments. Monitoring key product metrics. Understanding the root causes of changes in metrics. Building and analyzing dashboards and reports. Building key data sets to empower operational and exploratory analysis. Evaluating and defining metrics

Exploratory Analysis: Proposing what to build in the next roadmap. Understanding ecosystems, user behaviors, and long-term trends. Identifying new levers to help move key metrics. Building models of user behaviors for analysis or to power production systems

Product Leadership: Influencing product teams through the presentation of data-based recommendations. Communicating state of business, experiment results, etc. to product teams. Spreading best practices to analytics and product teams

Data Infrastructure: Working in Hadoop and Hive primarily, sometimes MySQL, Oracle, and Vertica. Automating analyses and authoring pipelines via SQL and Python-based ETL framework

For additional insights, refer to the Roblox Data Scientist guide.

Meta Data Scientist Product analytics Interview Guide

The initial interview will be a 45-minute video conference with a Meta data scientist.

The interview will include questions and discussion around both product interpretation

and applied data, as well as a few minutes for your questions at the end.

Skills the interviewer is looking for:

Framing: Can you structure and see data to answer a fairly open-ended question?

Operationalization: Can you translate the concepts generated into specific actions?

Analytical Understanding: Can you translate between numbers and words (i.e. prove to your interviewer that product “X” should be built through data resulting in analytical proof)?

Hypothesis Driven: Can you identify reasonable hypotheses and apply basic logic to support those hypotheses? Can you identify hypotheses, and do you understand how to look at data to confirm or refute a product insight?

As mentioned earlier the initial screener has questions from 2 parts which are:

Product Interpretation-

This part of the interview is a product case study focused on interpreting user behavior using data and metrics. It focuses broadly on how you translate user behavior into product ideas and insights using data and metrics. A sample question might be positioned as: “How would you evaluate YouTube’s video recommendations?”The interviewer will be assessing your ability to:

Understand hypotheses for launching new features: “How can I improve a product?”

Consider and quantify tradeoffs of a feature in terms of metrics.

Design experiments to test these hypotheses.

Interpret results of experiments.

Communicate decision-making via metrics.

Applied Data-

The applied data part of your interview focuses more on the technical side of solving a problem

using data, for example: “How do you frame a problem, from selecting the most suitable data sets all the way down to execution?” Or, “How would you evaluate YouTube’s video recommendations?”

It would be worth your while to go through Meta's core products and also engage with each of their core products, trying to reverse-engineer in your mind how these products came to be, what metrics, and what testing and experimentation were involved.

Interview tips:

1.Think out loud.

Narrate your approach to the problem/question asked as you go through the problem so that the interviewer has insight into your thought process.

2. Deconstruct problems.

Follow the modular thinking approach to big ambiguous problems, breaking them into smaller groups, and combining the groups for a solution.

3. Hints.

Resort to mid answer course correction if your interviewer prompts you that you're heading in the wrong direction.

4. Clarification.

Ask clarifying questions during the interview.

5. Prepare an answer to the cliched "Why Meta?" question.

Meta interviewers like to see people who know about the company's environment, projects, challenges, etc.

6. Questions.

If time permits you may pop in a few questions yourself, say about Meta and analytics.

The Meta DS Product Analytics interview might be a challenging one to crack, but if your preparation is on the lines of the guide we have prepared, we believe you are definitely going to come off with flying colors.

Technical Skills & Utilities (TSU)

CORE TOPICS

Coding, Data Engineering, & Testing

BROAD TOPIC

SUB-TOPICS

DETAILED TOPICS

SQL

SELECT, JOIN

Window Functions, CTEs, Recrusions

Python

Pandas, Numpy

...

Statistics

[Refer to: Statistics and Mathematics (SMS)]

Machine Learning

Scikit-Learn

XGBoost, Deep Learning, Feature Stores

Data Engineering

Airflow

Dragster (Prefect), dbt, macros, CI/CD

Visualization

Tableau, Looker, Amplitude

Custom d3.js, Dashboard Parameters (Filtering)

Cloud & DevOps

BigQuery, AWS

Terraform, Kubernetes

Experimentation

T-Tests

Sequential Testing, CUPED, Heterogeneous-Effect Models

Layer

2025-leading Options

Warehouses / Lakehouses

BigQuery, Snowflake, Databricks, Microsoft Fabric OneLake

Batch & Stream Ingest

Fivetran, Airbyte, Kafka, Google Dataflow

Transformation (ELT)

dbt, pyspark-on-Databricks, SQLMesh

Orchestration

Apache Airflow, Dagster, Prefect

Languages

SQL, Python, R, Julia; Polars & DuckDB for local analytics

BI / Viz

Looker, Power BI, Tableau, Superset; streamlit/Plotly for apps

ML & MLOps

scikit-learn, XGBoost, PyTorch/TensorFlow, MLflow, Vertex AI

Experimentation Platforms

Optimizely, Eppo, VWO, in-house CUPED frameworks

Reverse-ETL / Activation

Census, Hightouch, RudderStack

Data Observability

Monte Carlo, Soda, Databand

Collaboration & Versioning

GitHub, GitLab, JupyterLab, Hex, Deepnote

Competency

Skill Level & Key Concepts

Core Tools

Application in a Project (Use Case)

SQL

Foundational: SELECT, WHERE, GROUP BY, ORDER BY, INNER/LEFT JOINs.
 Advanced: Window Functions (ROW_NUMBER, LEAD), Common Table Expressions (CTEs), Subqueries, performance tuning.

PostgreSQL, MySQL, SQL Server, BigQuery, Redshift

Extracting, segmenting, and aggregating customer subscription and product usage data from the company's main database.

Spreadsheets

Foundational: VLOOKUP/XLOOKUP, INDEX-MATCH, Pivot Tables, data filters.
 Advanced: Power Query (Get & Transform), creating data models, complex nested formulas.

Excel, Google Sheets

Performing quick, ad-hoc analysis on a smaller data export; validating data quality; creating simple charts for a team meeting.

Python

Foundational: Data structures, functions, loops, using Jupyter Notebooks.

Advanced: Mastery of key libraries:

- Pandas: DataFrames for cleaning, transforming, merging, & grouping.
- NumPy: Numerical operations.
- Matplotlib/Seaborn: Data visualization.
- Scikit-learn: Building predictive models (e.g., regression).
- Statsmodels: Rigorous statistical testing.

Jupyter, VS Code, Python, Pandas, Scikit-learn

Writing a script to clean a 500k-row dataset, perform exploratory data analysis, and build a logistic regression model to predict customer churn.

BI Platforms

Foundational: Connecting to data sources (CSVs, databases), building standard charts (bar, line, scatter).

Advanced: Creating interactive dashboards with filters and actions, using calculated fields, data storytelling.

Tableau, Power BI, Looker Studio

Creating a C-level interactive dashboard that reports on key churn drivers and allows stakeholders to drill down into specific customer segments.

Statistics

Foundational: Descriptive Statistics (mean, median, standard deviation), basic probability.

Advanced: Inferential Statistics (Hypothesis/A/B testing, t-tests), Regression Analysis (Linear, Logistic).

Python (Statsmodels, Scikit-learn), Excel

Designing and analyzing an A/B test to determine if a new app feature significantly reduced customer churn compared to the old version.

Top Technologies & Tools by Frequency

Rank

Technology/Tool

Count

Category

1

SQL

27

Programming Language

2

Python

23

Programming Language

3

R

11

Programming Language

4

Tableau

10

Data Visualization

5

Looker

9

Business Intelligence

6

dbt/DBT

14

Data Engineering

7

AWS

5

Cloud Platform

8

Salesforce

5

CRM Platform

9

Excel

5

Spreadsheet Tool

10

DOMO

4

Analytics Platform

11

Power BI

3

Business Intelligence

12

Snowflake

3

Cloud Data Platform

13

Databricks

3

Analytics Platform

14

BigQuery

3

Cloud Database

15

Google Analytics

3

Web Analytics

SQL (27 mentions) - Most critical technical skill

Python (23 mentions) - Primary programming language

R (11 mentions) - Statistical analysis focus

Tableau (10 mentions) - Leading visualization platform

dbt (14 combined mentions) - Modern data transformation

Modern Data Stack:

Cloud-First: AWS (5), Snowflake (3), BigQuery (3), Databricks (3)

Data Engineering: dbt (14), Airflow (1), ETL (2)

Self-Service Analytics: Looker (9), Tableau (10), Power BI (3)

Programming-Heavy: Python (23), R (11), SQL (27)

Traditional vs Modern Tools:

Traditional: Excel (5), PowerPoint (implied)

Modern BI: Tableau (10), Looker (9), DOMO (4)

Cloud Analytics: Snowflake (3), BigQuery (3), Databricks (3)

Technical and Analytical Questions

Gemini

Tool Proficiency: What tools and technologies are you most comfortable with for marketing analytics? (e.g., Google Analytics, Tableau, SQL, Python, R)

Data Modeling: How do you design predictive models for customer lifetime value (CLV), churn, or lead scoring?

Attribution Models: Can you explain the different types of marketing attribution models and when to use each?

Data Quality: How do you ensure the accuracy and consistency of marketing data across multiple sources?

Advanced Analytics: Describe a situation where you used advanced statistical methods to solve a marketing problem.

Grok

Can you explain how you would measure the ROI of a marketing campaign?

This tests your understanding of key performance indicators (KPIs), attribution models, and financial metrics.

What analytics tools are you proficient in?

Expect to discuss software like Google Analytics, Adobe Analytics, Tableau, SQL, Python, or R for data manipulation and visualization.

How do you ensure data quality in your analytics work?

This can involve discussing data cleaning processes, validation checks, and maintaining data integrity.

Describe an instance where you used predictive analytics in marketing.

They'll want to see your ability to use data to forecast trends or consumer behavior.