Technical Interview Questions for Data Tracks at ITI

Answers of questions in this link
https://datapost73.blogspot.com/2020/08/iti-40-datamanagement-data-science.html?fbclid=IwAR0e6VCxxcNVzaEMuaCqP0KVSJJIkFl9Ob9JuwGtEwhxbDgRYrz_IcwN240

Index	Question	Answer
SQL + PL / SQL + Database (Very Important topic)		
1	What are DCL, DML, and DDL in SQL?	DCL (Data Control Language) is used to manage permissions and access control. DML (Data Manipulation Language) is used for data manipulation like INSERT, UPDATE, DELETE. DDL (Data Definition Language) is used to define and manage database structures like CREATE, ALTER, DROP. Example: DCL - GRANT SELECT ON table TO user; DML - INSERT INTO table (column1, column2) VALUES (value1, value2); DDL - CREATE TABLE table (column1 datatype, column2 datatype);
2	What is the difference between group by and having?	GROUP BY is used to group rows based on a column's values, typically used with aggregate functions. HAVING is used to filter grouped results. Example: SELECT department, AVG(salary) FROM employees GROUP BY department HAVING AVG(salary) > 50000;
3	What is the order by? Can we order more than one column?	ORDER BY is used to sort query results. You can order by one or more columns by specifying multiple column names in the ORDER BY clause. Example: SELECT name, age FROM students ORDER BY age, name;
4	What is the difference between union and join?	UNION combines the result sets of two or more SELECT queries into a single result set, removing duplicates. JOIN combines rows from two or more tables based on a related column. Example: UNION - SELECT name FROM table1 UNION SELECT name FROM table2; JOIN - SELECT customers.name, orders.order_date FROM customers JOIN orders ON customers.customer_id = orders.customer_id;
5	What is the different type of join?	There are various types of joins: INNER JOIN (returns matching rows), LEFT JOIN (returns all rows from the left table and matching rows from the right), RIGHT JOIN (returns all rows from the right table and matching rows from the left), FULL OUTER JOIN (returns all rows when there is a match in either table). Example: INNER JOIN - SELECT customers.name, orders.order_date FROM customers INNER JOIN orders ON customers.customer_id = orders.customer_id;
6	What is the aggregate functions?	Aggregate functions perform calculations on a set of values and return a single result. Common aggregates include COUNT, SUM, AVG, MAX, and MIN. Example: SELECT COUNT(*) FROM orders;
7	What are the SQL statements Sequence?	The typical sequence of SQL statements in a query is: SELECT (columns) FROM (table)

		WHERE (conditions) GROUP BY (columns) HAVING (conditions) ORDER BY (columns);
8	What is the view? + Why we use it?	A view is a virtual table based on the result of a SELECT query. It simplifies complex queries, provides security, and hides underlying table structures. Example: CREATE VIEW employee_view AS SELECT name, salary FROM employees WHERE department = 'HR';
9	What is the SQL transaction?	A SQL transaction is a sequence of one or more SQL statements treated as a single unit of work. It follows ACID properties (Atomicity, Consistency, Isolation, Durability) to ensure data integrity. Example: BEGIN TRANSACTION; UPDATE account SET balance = balance - 100 WHERE account_number = '123'; COMMIT;
10	What is the difference between delete and truncate?	DELETE removes specific rows from a table based on a condition and can be rolled back. TRUNCATE removes all rows from a table and is not reversible. Example: DELETE FROM employees WHERE department = 'IT'; TRUNCATE TABLE employees;
11	How can insert a column to the table?	You can use the ALTER TABLE statement to add a column to an existing table. Example: ALTER TABLE employees ADD COLUMN address VARCHAR(255);
12	How can insert multi rows in only one insert statement?	Use the INSERT INTO statement with multiple value sets in parentheses. Example: INSERT INTO students (name, age) VALUES ('Alice', 25), ('Bob', 22), ('Charlie', 28);
13	What is the database, DBMS, and RDBMS?	A database is a structured collection of data. DBMS (Database Management System) is software that manages databases. RDBMS (Relational DBMS) stores data in tables with relationships. Example: Database: CompanyDB; DBMS: MySQL; RDBMS: PostgreSQL;
14	What are the kinds of attributes?	Attributes in a database represent properties of entities. They can be classified as simple (atomic) or composite (composed of sub-attributes) and derived (calculated from other attributes). Example: Simple - Age, Composite - Address (Street, City), Derived - TotalPrice (Quantity * Price);
15	What is the ERD?	ERD (Entity-Relationship Diagram) is a visual representation of database entities, their attributes, and relationships between entities. It helps in database design. Example: ;
16	What is the type of constraints?	Constraints enforce data integrity rules. Common types include PRIMARY KEY, FOREIGN KEY, UNIQUE, CHECK, and NOT NULL. Example: PRIMARY KEY (employee_id), FOREIGN KEY (department_id) REFERENCES departments(department_id);
17	What is the difference between primary key and foreign key?	A primary key uniquely identifies rows in a table. A foreign key establishes a link between tables, ensuring referential integrity. Example: PRIMARY KEY - employee_id in employees table; FOREIGN KEY - department_id in employees table referencing departments table;
18	What the difference is between delete and truncate?	(Repeated question) DELETE removes specific rows; TRUNCATE removes all rows and is not reversible.
19	What is delete set null and delete cascade?	DELETE SET NULL sets foreign key values to NULL when referenced rows are deleted. DELETE CASCADE deletes rows in related tables when the referenced row is deleted. Example: DELETE SET

		NULL - Set employee_id to NULL in orders when an employee is deleted; DELETE CASCADE - Delete all orders when an employee is deleted.
20	What is the normalization and why are we making it?	Normalization is the process of organizing data in a database to reduce redundancy and improve data integrity. It prevents update anomalies and ensures efficient data storage. Example: 1NF - Ensure each column has atomic values; 2NF - Remove partial dependencies; 3NF - Remove transitive dependencies.
21	What are the types of normalization?	Common normalization forms include 1NF, 2NF, 3NF, BCNF, and 4NF. Each eliminates specific types of data redundancy. Example: 1NF - Each column has atomic values; 2NF - No partial dependencies; 3NF - No transitive dependencies.
22	What are the update anomalies?	Update anomalies occur when inconsistencies arise due to data redundancy, such as when updating data in one place but not another. Example: In a denormalized table, updating an employee's salary in one row but not in another for the same employee.
23	What is the difference between SQL and PL/SQL?	SQL is a query language for managing and querying data in databases. PL/SQL is a procedural extension of SQL used for writing stored procedures and functions. Example SQL: SELECT * FROM employees; Example PL/SQL: CREATE PROCEDURE getEmployee (emp_id NUMBER) AS BEGIN ... END;
24	What are the types of loops in PL/SQL?	PL/SQL provides loops like FOR LOOP, WHILE LOOP, and LOOP-END LOOP for repetitive tasks. Example: FOR i IN 1..10 LOOP ... END LOOP;
25	What are the cursors and what are the cursors types?	Cursors are database objects used to retrieve and manipulate data. Types include Implicit (used for single-row queries) and Explicit (used for multi-row queries), which can be further categorized as Static, Dynamic, and Scrollable. Example: Implicit Cursor - SELECT name INTO employee_name FROM employees WHERE id = 123; Explicit Cursor - DECLARE emp_cursor CURSOR FOR SELECT name FROM employees;
26	What is the procedure?	A procedure is a named collection of PL/SQL statements that can be stored in a database and executed as a single unit. It can take parameters and return values. Example: CREATE PROCEDURE calculate_salary (employee_id NUMBER) AS BEGIN ... END;
27	What is the difference between procedure and function?	A procedure doesn't return a value, while a function does. Functions can be used in SQL queries, whereas procedures cannot. Example Procedure: CREATE PROCEDURE update_employee (emp_id NUMBER) AS BEGIN ... END; Example Function: CREATE FUNCTION get_employee_name (emp_id NUMBER) RETURN VARCHAR2 AS BEGIN ... END;
28	What are the triggers and what are the triggers types?	Triggers are PL/SQL blocks executed automatically in response to specific database events. Types include BEFORE and AFTER triggers for INSERT, UPDATE, DELETE events. Example: BEFORE INSERT Trigger - Prevent inserting records with invalid data; AFTER UPDATE Trigger - Log changes to a table.
29	Write SQL Statements	SQL statements depend on specific requirements and tables. For example, to insert data: INSERT INTO employees (emp_id, emp_name) VALUES

		(1, 'John Doe'); To update data: UPDATE products SET price = price * 0.9 WHERE category = 'Electronics'; To delete data: DELETE FROM customers WHERE last_purchase_date < '2022-01-01';
Business Intelligence		
30	What is Business Intelligence?	Business Intelligence (BI) refers to the technologies, processes, and tools used to analyze and present business data to support decision-making. It helps organizations gain insights, make informed decisions, and improve business performance. Example: Using BI to analyze sales data to identify trends and optimize product offerings.
31	What are the steps in BI?	The typical steps in BI include data collection, data integration (ETL - Extract, Transform, Load), data storage, data analysis, and data visualization. Example: 1. Collecting sales data from multiple sources. 2. Integrating and transforming the data into a unified format. 3. Storing it in a data warehouse. 4. Analyzing it to discover sales trends. 5. Creating dashboards to visualize the trends.
32	What are the tools we use in BI (for ETL, Analysis, and Visualization)?	ETL (Extract, Transform, Load) tools include Apache NiFi, Talend, and Informatica. Analysis tools include Tableau, Power BI, and QlikView. Visualization tools include D3.js, Google Data Studio, and Looker. Example: Using Tableau for data analysis and visualization to create interactive sales reports.
Data Warehouse		
33	What is the data warehouse?	A data warehouse is a centralized repository that stores, integrates, and manages data from various sources to support business reporting and analysis. It is designed for query and analysis rather than transaction processing. Example: Storing historical sales data, customer information, and product data for business intelligence purposes.
34	What are the characteristics of a data warehouse?	Characteristics include subject-oriented (focus on specific business areas), integrated (combines data from diverse sources), time-variant (stores historical data), non-volatile (data is not updated frequently), and supports complex queries. Example: Analyzing sales trends over the last five years.
35	What is the difference between a database and a data warehouse?	A database is designed for transactional processing, while a data warehouse is designed for analytical processing. Databases support real-time data updates, while data warehouses store historical data and support complex queries for reporting and analysis. Example: A database for online order processing vs. a data warehouse for sales analysis.
36	What is the difference between a data warehouse and big data?	A data warehouse stores structured data in a highly organized manner, while big data encompasses vast volumes of structured and unstructured data. Data warehouses are well-suited for structured data analysis, whereas big data technologies like Hadoop handle unstructured and semi-structured data. Example: A data warehouse for analyzing sales data vs. using big data tools to analyze social media posts.

37	What is the difference between OLTP and OLAP?	OLTP (Online Transaction Processing) systems are used for day-to-day transactional operations, supporting real-time data entry and retrieval. OLAP (Online Analytical Processing) systems are for complex data analysis and reporting. Example: OLTP for processing bank transactions, OLAP for analyzing customer spending patterns.
38	What is Data Warehousing?	Data warehousing is the process of designing, building, and maintaining data warehouses. It involves data extraction, transformation, loading (ETL), and providing a platform for business intelligence and reporting. Example: Setting up a data warehousing system for a retail company.
39	What are the processes that can be done in the data warehouse?	Processes include data extraction, data transformation, data loading (ETL), data storage, data retrieval, data modeling, and data analysis. Example: Extracting customer data, transforming it into a standardized format, and loading it into the data warehouse for analysis.
40	What is Data Modeling? + Types of Data Modeling?	Data modeling is the process of defining the structure and relationships of data in a database or data warehouse. Types include conceptual modeling (high-level representation), logical modeling (entity-relationship diagrams), and physical modeling (designing database tables). Example: Creating an entity-relationship diagram for a customer database.
41	Can we update a record in a data warehouse?	Data warehouses are typically designed for read-intensive operations, and updates are infrequent. Updates can be performed, but they often involve complex ETL processes to maintain historical data. Example: Correcting a customer's address in the data warehouse.
42	What is a data mart?	A data mart is a subset of a data warehouse that focuses on specific business areas or departments. It contains a smaller, more specialized set of data for targeted analysis. Example: Creating a sales data mart for the Sales department to analyze sales performance.
43	What is a Data Cube?	A data cube is a multi-dimensional representation of data that allows for efficient querying and analysis. It contains dimensions (attributes) and measures (facts) and is often used in OLAP systems. Example: Analyzing sales data with dimensions like time, product, and region.
44	What is ETL?	ETL (Extract, Transform, Load) is a process used to extract data from source systems, transform it into a desired format, and load it into a data warehouse or data mart. Example: Extracting sales data from a CRM system, transforming it to match the data warehouse schema, and loading it into the data warehouse.
45	What is the difference between snowflake and star schema?	In a star schema, dimension tables are directly linked to a central fact table. In a snowflake schema, dimension tables are normalized into multiple related tables. Star schemas are simpler but can be less space-efficient, while snowflake schemas save space but can be more complex. Example: Star schema for sales analysis vs. snowflake schema for complex product hierarchies.
46	What is the difference between fact and dimension tables?	Fact tables contain numerical measures and foreign keys to dimension tables. Dimension tables contain descriptive attributes about dimensions

		such as time, product, or location. Example: Fact table with sales revenue vs. dimension table with product details.
	Big Data	
47	Why is big data important?	Big data is important because it enables organizations to gain valuable insights from vast and diverse datasets that were previously too large and complex to manage and analyze effectively. It can uncover patterns, trends, and opportunities for better decision-making. Example: Analyzing customer behavior across social media, online purchases, and offline interactions to enhance marketing strategies.
48	What is big data? (V's of Big Data)	Big data is characterized by the three V's: Volume (large amounts of data), Velocity (high-speed data generation and processing), and Variety (diverse data types, structured and unstructured). Some also add Veracity (data accuracy) and Value (extracting insights). Example: Social media platforms processing massive volumes of tweets (Volume) in real-time (Velocity) with text, images, and videos (Variety).
49	What are the data types?	Data types in the context of big data can include structured data (e.g., numbers, dates), semi-structured data (e.g., JSON, XML), and unstructured data (e.g., text, images, videos). Example: Structured data - Sales revenue as numbers; Semi-structured data - Customer data in JSON format; Unstructured data - Text reviews from customers.
50	What is Data Lake?	A Data Lake is a central repository that stores vast amounts of raw and unprocessed data from diverse sources. It allows for flexible and scalable data storage and analysis. Example: Storing log files, sensor data, and social media posts in a Data Lake for future analytics.
51	What is the difference between ETL & ELT?	ETL (Extract, Transform, Load) involves extracting data from source systems, transforming it before loading it into a data warehouse. ELT (Extract, Load, Transform) loads data into the data warehouse first and then performs transformations. ELT is often used in big data scenarios where data may not fit the traditional ETL model. Example (ETL): Extracting sales data, aggregating it, and loading it into a data warehouse. Example (ELT): Loading raw log data into a Data Lake, then transforming it into a structured format for analysis.
52	What is the difference between a Database and Big data?	Databases are designed for structured data storage and transaction processing, while big data encompasses both structured and unstructured data. Big data technologies like Hadoop and NoSQL databases are built to handle massive volumes and varieties of data. Example: A relational database for storing customer information vs. Hadoop for processing social media data.
53	What are the tools in big data?	Big data tools include Hadoop (for distributed storage and processing), Spark (for fast data processing), MapReduce (for data processing in Hadoop), Hive (for querying and data warehousing), Impala (for SQL queries on Hadoop), Kafka (for real-time data streaming), and

		more. Example: Using Spark for analyzing large datasets in real-time.
54	What are (Hadoop/Spark/MapReduce/Hive/Impala/Kafka/...)?	- Hadoop is a distributed storage and processing framework for big data. - Spark is a fast and versatile data processing engine. - MapReduce is a programming model used in Hadoop for parallel processing. - Hive is a data warehousing and SQL querying tool for Hadoop. - Impala is an open-source SQL query engine for Hadoop. - Kafka is a distributed streaming platform for real-time data. Example: Using Hadoop to store and process large log files, Spark for real-time analytics, Hive for querying structured data in Hadoop, and Kafka for ingesting streaming data.
Data Science + Machine Learning + Data Mining (Data Science Track)		
55	What is data science?	Data science is an interdisciplinary field that uses scientific methods, algorithms, processes, and systems to extract knowledge and insights from structured and unstructured data. It combines aspects of statistics, computer science, and domain knowledge to solve complex problems. Example: Using data science to analyze customer behavior and recommend personalized products.
56	What is the difference between data scientists and data analysts?	Data scientists focus on designing and implementing complex algorithms to solve business problems, often requiring programming and machine learning expertise. Data analysts primarily work on data exploration, visualization, and basic statistical analysis to answer specific questions. Example: A data scientist develops a predictive model, while a data analyst creates reports and dashboards.
57	What is data cleaning? How do we clean the data?	Data cleaning involves identifying and correcting errors, inconsistencies, and inaccuracies in datasets. It includes tasks like handling missing values, removing duplicates, and correcting outliers using statistical methods and domain knowledge. Example: Replacing missing age values in a dataset with the median age of known values.
58	What is Data Mining?	Data mining is the process of discovering patterns, relationships, and valuable insights from large datasets. It involves techniques like clustering, classification, regression, and association rule mining. Example: Analyzing retail sales data to identify product associations for marketing strategies.
59	What are the real-life applications of data mining and machine learning?	Applications include fraud detection, recommendation systems (e.g., Netflix), medical diagnosis, sentiment analysis in social media, predictive maintenance in manufacturing, and autonomous vehicles. Example: Using machine learning to predict disease outbreaks based on historical health data.
60	What is the Process of Data Mining/Knowledge Discovery Process?	The process involves data selection, data preprocessing, data transformation, data mining, pattern evaluation, and knowledge presentation. Example: In e-commerce, selecting sales data, preprocessing it (cleaning and transforming), mining customer purchase patterns, and presenting these patterns for business decisions.
61	What are the Challenges of Data Mining?	Challenges include handling large datasets, data quality issues, selecting appropriate algorithms,

		overfitting, interpretability of complex models, and ensuring privacy and security of sensitive data. Example: Dealing with skewed data distribution in fraud detection, where fraudulent transactions are rare.
62	What is Machine Learning?	Machine learning is a subset of artificial intelligence that involves the development of algorithms that enable computers to learn patterns and make predictions or decisions from data. Example: Training a machine learning model to recognize handwritten digits in images.
63	What is deep learning?	Deep learning is a subset of machine learning that uses artificial neural networks with multiple layers (deep architectures) to automatically learn and represent data. It excels in tasks like image and speech recognition. Example: Training a deep neural network to recognize objects in images.
64	What are the data mining tasks/algorithms?	Tasks include clustering (K-Means), classification (Decision Trees), regression (Linear Regression), association rule mining (Apriori), and anomaly detection (Isolation Forest). Example: Using K-Means to group customers based on purchasing behavior.
65	What is the difference between Supervised and Unsupervised learning?	Supervised learning uses labeled data to train models (e.g., classification or regression), while unsupervised learning uses unlabeled data to find patterns or groupings (e.g., clustering). Examples: Supervised - Spam email detection; Unsupervised - Customer segmentation.
66	What is the difference between Classification and Clustering?	Classification assigns labels to data based on predefined classes, while clustering groups data into clusters based on similarity. Examples: Classification - Identifying email as spam or not; Clustering - Grouping customers into market segments.
67	Examples of clustering algorithms	K-Means, Hierarchical Clustering, and DBSCAN are examples of clustering algorithms.
68	Examples for classification algorithms	Decision Trees, Logistic Regression, and Support Vector Machines (SVM) are examples of classification algorithms.
69	What is an association rule?	Association rules identify relationships between items in a dataset, often used in market basket analysis to find item associations in transactions. Example: "If a customer buys bread, they are likely to buy butter."
70	How does this algorithm work (K-Mean, Regression, SVM, association rule, decision tree, KNN...)?	Provide brief explanations of how each algorithm works. Example: K-Means clusters data points into K clusters based on proximity; Decision Trees make decisions by following a tree-like structure of if-else conditions.
71	What is recall and precision, F1?	Recall measures the ability of a model to identify all relevant instances. Precision measures the ability of a model to return only relevant instances. F1-score is the harmonic mean of precision and recall, balancing them. Example: In a medical test, recall is the percentage of actual sick patients correctly identified by the test.
72	What is the bias-variance trade-off?	The bias-variance trade-off refers to the balance between model complexity and model performance. A model with high bias (underfitting) has low complexity and may not capture underlying patterns. A model with high variance (overfitting) fits the training data too closely and may not generalize well to new data. Example: In

		polynomial regression, increasing the polynomial degree leads to lower bias but higher variance.
73	What is the confusion matrix?	A confusion matrix is a table that visualizes the performance of a classification algorithm. It shows true positives, true negatives, false positives, and false negatives. Example: In a binary classification problem, the confusion matrix may look like this: TP: 120, TN: 80, FP: 10, FN: 5.
74	What is the ROC Curve?	The ROC (Receiver Operating Characteristic) curve is a graphical representation of a classifier's performance, showing the trade-off between true positive rate and false positive rate at various thresholds. Example: In medical diagnosis, plotting the ROC curve helps assess the accuracy of a diagnostic test.
75	Explain cross-validation?	Cross-validation is a technique used to evaluate the performance of a machine learning model by dividing the dataset into multiple subsets (folds). It trains and tests the model on different combinations of folds to assess its generalization ability. Example: Using k-fold cross-validation to train and test a model on five subsets of the data, rotating which subset is used for testing in each iteration.
76	What is the difference between a validation set and a test set?	A validation set is used during the model training phase to tune hyperparameters and assess performance. A test set is a separate dataset used to evaluate the final model's generalization performance. Example: Using a validation set to adjust the learning rate in gradient boosting and a test set to estimate the model's accuracy on unseen data.
77	How do you treat missing/outlier values?	Missing values can be imputed using methods like mean, median, or interpolation. Outliers can be identified and removed or transformed using statistical techniques. Example: Replacing missing age values with the median age of known values; Detecting outliers using the Z-score and removing extreme values.
78	How do you prepare the data for the ML Model?	Data preparation involves data cleaning, feature selection/engineering, handling missing values, scaling/normalizing features, and splitting data into training, validation, and test sets. Example: Scaling numerical features to have a mean of 0 and a standard deviation of 1 for better model convergence.
Statistics (Data Science Track)		
79	What is the difference between standard deviation and variance?	Variance measures how individual data points deviate from the mean. Standard deviation is the square root of the variance and measures the average deviation of data points from the mean. Example: Variance calculates the average squared difference from the mean, while standard deviation provides a more interpretable measure in the original units of the data.
80	What are Mean, Median, and Mode?	Mean is the average of a set of numbers. Median is the middle number when the numbers are ordered. Mode is the value that appears most frequently. Example: For the set of numbers {2, 3, 3, 5, 7}, Mean = 4, Median = 3, Mode = 3.
81	What is the difference between variance and standard deviation?	Variance measures the spread or dispersion of data by calculating the average of squared differences from the mean. Standard deviation is the square root of the variance and provides a

		more interpretable measure in the original units of the data. Example: Variance = 9, Standard Deviation = 3 for the set {1, 2, 3, 4, 5}.
82	What is the Box plot?	A box plot (box-and-whisker plot) is a graphical representation of the distribution of data. It shows the median, quartiles, and potential outliers. The box represents the interquartile range (IQR), and the whiskers extend to the minimum and maximum values within a defined range. Example: A box plot showing the distribution of test scores, with the median, quartiles, and any outliers.
83	What are the types of skewed data?	Skewed data can be positively skewed (right-skewed) where the tail extends to the right, or negatively skewed (left-skewed) where the tail extends to the left. Example: Positive skew in income distribution data due to a few high earners; Negative skew in test scores with many high scores.
84	What is the Z-score?	The Z-score (standard score) measures how many standard deviations a data point is from the mean. It standardizes data, making it possible to compare values from different datasets. Example: A Z-score of -1.5 indicates a data point is 1.5 standard deviations below the mean.
85	What is the P-value?	The P-value measures the evidence against a null hypothesis in hypothesis testing. It indicates the probability of observing a test statistic as extreme as, or more extreme than, what is observed in the sample, assuming the null hypothesis is true. Example: In a medical trial, a P-value of 0.03 suggests a 3% chance of observing the results if the treatment has no effect (null hypothesis).
86	What is the Pearson correlation coefficient?	The Pearson correlation coefficient (Pearson's r) measures the linear relationship between two continuous variables. It ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear correlation. Example: Pearson's r of 0.75 between hours studied and exam scores suggests a strong positive correlation.
87	What is A/B Testing?	A/B testing (split testing) is a controlled experiment where two versions (A and B) of a webpage, app, or product are compared to determine which performs better in terms of user engagement or conversions. Example: Testing two different website layouts to see which one results in higher click-through rates.
88	What is hypothesis testing?	Hypothesis testing is a statistical method used to make inferences about population parameters based on sample data. It involves formulating a null hypothesis (no effect) and an alternative hypothesis (an effect exists) and testing the null hypothesis using data and statistical tests. Example: Testing whether a new drug is more effective than an existing one in a clinical trial.