# Time Series Fundamentals: Questions and Answers

Time Series Forecasting  
What is time series forecasting?  
  
Predicting future values using past time-dependent data.  
What is the difference between a time series and a cross-sectional dataset?  
  
Time series: Observations over time.  
Cross-sectional: Data from multiple entities at one point in time.  
What are the key components of time series data, and why is it important to identify trends, seasonality, and noise?  
  
Trend (long-term direction), seasonality (repeating cycles), noise (random variation).  
Helps in accurate forecasting and model selection.  
How would you handle missing data or outliers in a time series dataset before applying a forecasting model?  
  
Missing data: Imputation (mean, forward fill, interpolation).  
Outliers: Detect using Z-score/IQR and handle via smoothing/capping.  
Can you describe the difference between univariate and multivariate time series forecasting? When would you prefer one over the other?  
  
Univariate: Uses one variable (e.g., past sales to predict future sales).  
Multivariate: Uses multiple variables (e.g., sales, weather, promotions).  
Use multivariate when external factors impact predictions.  
What does ARIMA stand for, and what are the key components of an ARIMA model?  
  
ARIMA = AutoRegressive (AR) + Integrated (I) + Moving Average (MA).  
AR uses past values, I makes data stationary, MA uses past errors.  
SQL  
What is a JOIN in SQL?  
  
Combines rows from two tables based on a common column.  
What is the difference between an INNER JOIN and a LEFT JOIN?  
  
INNER JOIN: Returns matching rows from both tables.  
LEFT JOIN: Returns all rows from the left table and matching right-table rows, filling missing values with NULL.  
Write an SQL query to join two tables using an INNER JOIN.  
  
SELECT e.name, d.department\_name   
FROM employees e   
INNER JOIN departments d ON e.department\_id = d.id;  
How do you perform a FULL OUTER JOIN in SQL?  
SELECT \*   
FROM employees   
FULL OUTER JOIN departments ON employees.department\_id = departments.id;  
(Some databases like MySQL do not support FULL OUTER JOIN directly.)  
  
Can you join a table with itself? What is it called?  
  
Yes, it’s called a self-join (e.g., finding employee-manager relationships).  
What are aggregate functions in SQL? Name a few.  
  
Functions that perform calculations on data: COUNT, SUM, AVG, MAX, MIN.  
How do you find the average value of a column in SQL?  
  
SELECT AVG(salary) FROM employees;  
Write an SQL query to count the number of records in a table.  
  
SELECT COUNT(\*) FROM employees;  
How do you group records in SQL?  
  
  
SELECT department, COUNT(\*)   
FROM employees   
GROUP BY department;  
What is the HAVING clause used for in SQL?  
  
Filters aggregated results after GROUP BY, e.g., departments with more than 10 employees:  
  
SELECT department, COUNT(\*)   
FROM employees   
GROUP BY department   
HAVING COUNT(\*) > 10;  
Machine Learning  
What is machine learning, and how does it differ from traditional programming?  
  
ML enables computers to learn patterns from data instead of following explicit rules.  
What is the difference between supervised and unsupervised learning?  
  
Supervised: Uses labeled data (e.g., spam detection).  
Unsupervised: Finds patterns in unlabeled data (e.g., customer segmentation).  
Name a few types of supervised learning algorithms.  
  
Regression: Linear Regression, Decision Trees.  
Classification: Logistic Regression, Random Forest, SVM.  
What is the role of data in machine learning?  
  
Quality data improves model accuracy; bad data leads to poor predictions.  
How do you evaluate a machine learning model?  
  
Regression: RMSE, MAE, R².  
Classification: Accuracy, Precision, Recall, F1-score, ROC.  
What is the confusion matrix, and what does it tell you?  
  
A table showing TP, FP, TN, FN for classification models.  
What is accuracy in machine learning, and when is it not a reliable metric?  
  
Accuracy = Correct Predictions / Total Predictions.  
Not reliable for imbalanced datasets (e.g., fraud detection).  
What is F1-score, and why is it important?  
  
Harmonic mean of precision & recall, useful for imbalanced data.  
How is the ROC curve used to evaluate models?  
  
Plots True Positive Rate (TPR) vs. False Positive Rate (FPR); AUC indicates model performance.  
What is the difference between precision and recall?  
  
Precision: TP / (TP + FP) → Focus on correctness.  
Recall: TP / (TP + FN) → Focus on capturing all relevant cases.  
Trade-off: High precision = fewer false positives, high recall = fewer false negatives.  
Time Series Fundamentals  
What are the key components of a time series?  
  
Trend, seasonality, cyclic patterns, and noise.  
What is the difference between stationary and non-stationary time series?  
  
Stationary: Constant mean & variance.  
Non-stationary: Changes over time, requiring transformation.  
What is ARIMA, and how does it work?  
  
ARIMA (AutoRegressive Integrated Moving Average) models time series with differencing, past values (AR), and past errors (MA).  
What is the role of seasonality in time series forecasting?  
  
Seasonal patterns repeat at regular intervals and must be accounted for in models.  
How do you handle missing values in time series data?  
  
Forward fill, interpolation, or model-based imputation.  
What are lag features, and how are they used in forecasting?  
  
Previous values used as inputs to predict future values.  
What are the common performance metrics used for time series forecasting?  
  
MAE, RMSE, MAPE, and R².  
How do you perform cross-validation for time series data?  
  
Use time series split (expanding window or rolling window).  
Time Series Data Handling  
How do you handle missing values in time series data?  
  
Forward fill, backward fill, mean imputation, or regression-based imputation.  
What are different methods for detecting and handling outliers in time series data?  
  
Z-score, IQR, rolling statistics, and smoothing techniques.  
Why is stationarity important in time series analysis?  
  
Many models assume stationarity; check using ADF or KPSS tests.  
What are techniques for resampling and aggregating time series data?  
  
Downsampling (daily → monthly), upsampling (hourly → minute-level), moving averages.  
How do you handle seasonality and trends in time series data?  
  
Differencing, decomposition, and seasonal adjustment techniques.  
Data Preprocessing  
What are the main steps in data preprocessing?  
  
Cleaning, transformation, feature engineering, and scaling.  
How do you handle missing data?  
  
Deletion, imputation (mean/median), or model-based methods.  
What is the difference between normalization and standardization?  
  
Normalization: Scales between 0-1.  
Standardization: Centers to mean 0, variance 1.  
How do you detect and handle outliers?  
  
Z-score, IQR, or box plots.  
What are key techniques in EDA?  
  
Summary stats, histograms, box plots, correlation analysis.  
How do you analyze relationships between categorical and numerical variables?  
  
Chi-square test, ANOVA, correlation plots.  
What is multicollinearity, and how do you detect it?  
  
High correlation between independent variables; detect via VIF.  
SQL Joins & Queries  
Difference between INNER JOIN, LEFT JOIN, RIGHT JOIN, and FULL JOIN?  
  
INNER JOIN: Matches both tables.  
LEFT JOIN: All from left + matches.  
RIGHT JOIN: All from right + matches.  
FULL JOIN: All records, NULL where no match.  
Find the second highest salary from an Employee table.  
  
  
SELECT MAX(salary)   
FROM employees   
WHERE salary < (SELECT MAX(salary) FROM employees);  
Difference between WHERE and HAVING in SQL?  
  
WHERE filters rows, HAVING filters groups after aggregation.  
Detect and delete duplicate records.  
  
  
DELETE FROM employees  
WHERE id NOT IN (  
 SELECT MIN(id) FROM employees GROUP BY name, salary  
);  
Difference between RANK(), DENSE\_RANK(), and ROW\_NUMBER()?  
  
RANK(): Skips ranks on ties.  
DENSE\_RANK(): No gaps.  
ROW\_NUMBER(): Unique rank.  
Retrieve top N records from a table.  
  
SELECT \* FROM employees ORDER BY salary DESC LIMIT 5;  
Deep Learning  
Difference between ANN, CNN, and RNN?  
  
ANN: General neural network.  
CNN: Image processing.  
RNN: Sequential data (e.g., text).  
What is the vanishing gradient problem?  
  
Gradients shrink, slowing learning in deep networks.  
What is transfer learning?  
  
Using pre-trained models for new tasks, reducing training time.  
Why are activation functions important?  
  
Introduce non-linearity to help model complex patterns.  
How does batch normalization help?  
  
Stabilizes learning, speeds up convergence.  
Difference between LSTM and GRU?  
  
Both handle sequences, but GRU is computationally simpler.  
How to handle overfitting in deep learning?  
  
Dropout, early stopping, L2 regularization.  
Python Basics  
Basic data types in Python?  
  
int, float, str, bool, list, tuple, dict, set.  
Check variable type in Python?  
  
type(variable)  
Difference between mutable and immutable data types?  
  
Mutable: Lists, dicts.  
Immutable: Tuples, strings.  
Convert string to integer?  
  
int("42")  
Python script to check even or odd number.  
  
num = int(input())  
print("Even" if num % 2 == 0 else "Odd")  
Modify elements in a tuple?  
  
Tuples are immutable; convert to list first.  
Can you add elements to a tuple?  
  
No, tuples are immutable.  
Find the maximum number in a list?  
  
  
max([1, 2, 3, 4])  
Functions & Loops in Python  
Purpose of functions in Python?  
  
Reusable code blocks for efficiency.  
\*\*Difference between \*args and kwargs?  
  
\*args for variable-length positional arguments.  
\*\*kwargs for variable-length keyword arguments.  
Check if a string is a palindrome.  
  
  
def is\_palindrome(s):  
 return s == s[::-1]  
Slice a string in Python.  
  
  
s = "Hello"  
print(s[1:4]) # "ell"