I. BASIC PYTHON CONCEPTS (Common Across All Domains)

#	Question	Expected Knowledge / Example
1	What are Python's key features for data analysis?	Easy syntax, open-source, libraries like **pandas**, **numpy**, **matplotlib**, **scikit-learn**.
2	Difference between a list, tuple, and dictionary.	List = mutable, Tuple = immutable, Dict = key-value pairs.
3	How do you read a CSV file in Python?	import pandas as pd; df = pd.read_csv('data.csv')
4	Explain what a DataFrame is.	Two-dimensional labeled data structure used in pandas for tabular data.
5	How to handle missing data?	df.fillna(value) or df.dropna().
6	Difference between loc and iloc.	loc = label-based indexing, iloc = integer-based indexing.
7	How do you find summary statistics in Python?	df.describe()
8	How do you merge/join datasets?	pd.merge(df1, df2, on='ID')
9	Explain lambda functions and list comprehensions.	One-line anonymous functions and concise list creation.
10	How to visualize data in Python?	Using **matplotlib** (plt.bar(), plt.plot()) or **seaborn** (sns.heatmap()).

II. PYTHON FOR DATA ANALYTICS / FINANCE / MARKETING

#	Question	Practical Illustration
1	How to calculate ROI or profit using Python?	df['ROI'] = (df['Revenue'] - df['Cost']) / df['Cost']
2	How to group data for summary analytics?	df.groupby('Region')['Sales'].sum()

3	How to identify top 5 customers by sales?	df.sort_values('Sales', ascending=False).head(5)
4	How do you calculate correlation between variables?	df.corr() → checks relationships between Spend & Revenue.
5	How to create a pivot-style summary?	pd.pivot_table(df, values='Revenue', index='Region', columns='Channel', aggfunc='sum')
6	What is the use of NumPy in analytics?	Fast numerical computation; e.g. np.mean(df['Sales']).
7	How do you detect outliers?	Using IQR method or Z-score from scipy.stats.
8	Explain how Python integrates with Power BI.	Via Python scripting visuals for advanced charts or ML model scoring inside Power BI.

III. BASIC MACHINE LEARNING QUESTIONS

Conceptual Questions

#	**Question**	**Answer Expectation**
1	What is Machine Learning?	Algorithms that enable systems to learn patterns from data.
2	Difference between Supervised and Unsupervised learning.	Supervised → labeled data (Regression, Classification). Unsupervised → unlabeled data (Clustering, PCA).
3	What is Regression?	Predicting continuous outcomes (e.g., price, revenue).
4	What is Classification?	Predicting categorical outcomes (e.g., churn yes/no).
5	What is Overfitting and how to prevent it?	Model fits training data too well → use **cross-validation**, **regularization**, **early stopping**.

6	What is Feature Scaling?	Normalizing data (e.g., StandardScaler in scikit-learn).
7	What are Evaluation Metrics?	Regression → MSE, RMSE, R^2; Classification → Accuracy, Precision, Recall, F1-score.
8	Explain Train-Test Split.	Dividing dataset (e.g., 80-20) to assess model generalization.
9	What are the main steps in a ML project?	Data collection → Cleaning → Feature selection → Model building → Evaluation → Deployment.
10	Which libraries are commonly used?	**pandas**, **numpy**, **matplotlib**, **scikit-learn**, **xgboost**.

Coding / Practical Questions

Task	**Example / Syntax**
Train-test split	from sklearn.model_selection import train_test_split; X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
Linear Regression	from sklearn.linear_model import LinearRegression; model = LinearRegression().fit(X_train, y_train)
Model Prediction	y_pred = model.predict(X_test)
Evaluation	from sklearn.metrics import r2_score; r2_score(y_test, y_pred)
Confusion Matrix	from sklearn.metrics import confusion_matrix; confusion_matrix(y_test, y_pred)

IV. DOMAIN-SPECIFIC QUESTIONS

A. For Finance Students

#	**Question**	**Expected Focus**
1	How can ML be applied in credit-risk modeling?	Logistic Regression to predict default probability.

2	What is time-series forecasting?	Predicting financial trends (e.g., stock prices, sales).
3	How would you detect financial fraud using ML?	Anomaly detection models, Isolation Forest, clustering.
4	Which regression model would you use for stock-price prediction?	Linear Regression or LSTM (advanced).
5	How would you predict loan approval likelihood?	Classification (Logistic Regression, Decision Tree).
6	What KPIs can ML help predict in Finance?	Profit growth, ROI, NPV, default probability, churn risk.
7	How do you evaluate the accuracy of a financial model?	RMSE or R^2 for regression; confusion matrix for classification.
8	What is the difference between correlation and causation in finance analytics?	Correlation ≠ Causation; must verify business logic.

B. For Marketing Students

#	**Question**	**Expected Focus**
1	How can ML improve marketing campaigns?	Predictive segmentation, customer churn prediction, lead scoring.
2	What is Customer Segmentation?	Unsupervised learning (K-Means clustering).
3	Explain how to predict customer churn.	Logistic Regression or Random Forest using churn labels.
4	How can sentiment analysis help in marketing?	NLP models (VADER, TextBlob) to gauge customer opinions.
5	How would you use Linear Regression in marketing?	Predicting ad spend vs sales (Marketing Mix Modeling).

6	Which KPIs can ML help forecast?	CTR, Conversion Rate, ROI, Retention Rate.
7	How do you evaluate marketing model performance?	Accuracy, Precision-Recall, AUC-ROC curve.
8	What is Recommendation System?	Collaborative Filtering to recommend products to customers.

V. CROSS-DOMAIN APPLICATION QUESTIONS (Protiviti Style)

Question	**What They're Testing**
"If your model shows 95 % accuracy, how do you know it's reliable?"	Understanding of **overfitting, bias, variance.**
"How would you explain your ML model to a non-technical CFO/CMO?"	Communication skill & business storytelling.
"What are the limitations of using ML in marketing/finance?"	Awareness of data quality, ethical use, explainability.
"Which model would you choose to forecast quarterly revenue?"	Time-series or Regression reasoning.
"How would you decide which features to include?"	Feature importance and domain knowledge.

VI. QUICK PREPARATION CHECKLIST

Area	**Core Topics**
Python Basics	Data types, loops, functions, pandas, visualization
EDA (Exploratory Data Analysis)	Missing values, correlation, outliers
ML Algorithms	Linear/Logistic Regression, Decision Tree, K-Means
Evaluation Metrics	Accuracy, RMSE, Precision-Recall

Domain KPIs	ROI, CAC, CLV, Profit Margin, Churn Rate
Business Communication	Explain models with real-world context

VII. Sample 5 Rapid-Fire Questions (Protiviti Interview Style)

1. Explain how Python helps a financial analyst make better predictions.
2. How would you use ML to reduce marketing costs while maintaining ROI?
3. What's the difference between regression and classification?
4. Which Python libraries are essential for ML dashboards in Power BI?
5. How would you justify using a linear model instead of a complex deep-learning one in business terms?

A

I. SQL BASICS — FOUNDATIONAL QUESTIONS

#	Question	Expected Explanation / Example
1	What is SQL?	SQL (Structured Query Language) is used to store, manage, and retrieve data from relational databases like MySQL, SQL Server, PostgreSQL.
2	What are the main types of SQL commands?	- **DDL:** CREATE, ALTER, DROP - **DML:** SELECT, INSERT, UPDATE, DELETE - **DCL:** GRANT, REVOKE - **TCL:** COMMIT, ROLLBACK
3	What is the difference between WHERE and HAVING?	WHERE filters rows **before** aggregation; HAVING filters **after** aggregation. Example: SELECT Region, SUM(Sales) FROM Orders GROUP BY Region HAVING SUM(Sales)>50000;
4	Explain GROUP BY with example.	Used to group rows that share common values. SELECT Region, SUM(Revenue) FROM Sales GROUP BY Region;

5	What is the purpose of ORDER BY?	Sorts data ascending or descending. SELECT * FROM Sales ORDER BY Revenue DESC;
6	What are primary and foreign keys?	Primary key → unique identifier in a table. Foreign key → links data between tables.
7	Difference between INNER JOIN, LEFT JOIN, and FULL JOIN.	- **INNER JOIN:** Matching rows only - **LEFT JOIN:** All from left + matched from right - **FULL JOIN:** All records from both tables
8	Write SQL to find duplicate records.	SELECT Name, COUNT(*) FROM Customers GROUP BY Name HAVING COUNT(*)>1;
9	What is the use of DISTINCT?	Removes duplicate values. SELECT DISTINCT Region FROM Sales;
10	How do you find total, average, min, and max values?	SELECT SUM(Sales), AVG(Sales), MIN(Sales), MAX(Sales) FROM Orders;

II. INTERMEDIATE SQL CONCEPTS

#	**Question**	**Example / Explanation**
1	What is a subquery?	A query inside another query. SELECT * FROM Employees WHERE Salary > (SELECT AVG(Salary) FROM Employees);
2	What is a CTE (Common Table Expression)?	Temporary named result set. WITH RegionSales AS (SELECT Region, SUM(Revenue) AS Total FROM Sales GROUP BY Region) SELECT * FROM RegionSales WHERE Total>50000;
3	What is the difference between UNION and UNION ALL?	UNION removes duplicates, UNION ALL keeps them.
4	Explain window functions.	Functions like ROW_NUMBER(), RANK(), SUM() OVER() used without collapsing data.

		SELECT Region, SUM(Sales) OVER(PARTITION BY Region) AS Total FROM Sales;
5	How can you find the second-highest salary or sale?	SELECT MAX(Sales) FROM Sales WHERE Sales < (SELECT MAX(Sales) FROM Sales);
6	Explain normalization.	Process of organizing data to reduce redundancy (1NF, 2NF, 3NF).
7	What is a view?	Virtual table created using a query. CREATE VIEW HighValueCustomers AS SELECT * FROM Customers WHERE Revenue>100000;
8	How do you delete duplicate rows?	Use CTE + ROW_NUMBER(): sql WITH Temp AS (SELECT *, ROW_NUMBER() OVER(PARTITION BY Name ORDER BY Name) AS rn FROM Customers) DELETE FROM Temp WHERE rn>1;
9	Explain indexes.	Improve query speed on large tables (created on frequently searched columns).
10	How do you handle NULL values?	IS NULL, IS NOT NULL, or replace using COALESCE(Column, 0)

III. SQL FOR FINANCE ANALYTICS (Protiviti-Oriented)

#	**Question / Use Case**	**Sample Query / Concept**
1	Calculate total revenue and profit.	SELECT SUM(Revenue) AS TotalRev, SUM(Revenue - Cost) AS Profit FROM Finance;
2	Calculate profit margin by region.	SELECT Region, SUM(Revenue - Cost)/SUM(Revenue) AS ProfitMargin FROM Finance GROUP BY Region;
3	Show monthly sales trend.	SELECT MONTH(Date) AS Month, SUM(Revenue) AS Sales FROM Finance GROUP BY MONTH(Date) ORDER BY Month;

4	Find customers with overdue payments > 30 days.	SELECT * FROM Payments WHERE DATEDIFF(DAY, DueDate, PaymentDate) > 30;
5	Calculate Return on Investment (ROI).	SELECT (SUM(Revenue) - SUM(Investment))/SUM(Investment) AS ROI FROM Projects;
6	Identify top 5 highest revenue clients.	SELECT TOP 5 Client, SUM(Revenue) AS Total FROM Finance GROUP BY Client ORDER BY Total DESC;
7	Compare revenue of two financial years.	SELECT YEAR(Date), SUM(Revenue) FROM Finance WHERE YEAR(Date) IN (2023,2024) GROUP BY YEAR(Date);
8	Detect unusual transactions (fraud).	Use HAVING to filter high-value outliers: SELECT AccountID, SUM(Amount) FROM Transactions GROUP BY AccountID HAVING SUM(Amount) > 1000000;
9	Calculate average loan interest rate per region.	SELECT Region, AVG(InterestRate) FROM Loans GROUP BY Region;
10	Join balance sheet and income statement tables for analysis.	SELECT b.Account, i.Revenue, b.Assets FROM BalanceSheet b JOIN IncomeStatement i ON b.AccountID=i.AccountID;

IV. SQL FOR MARKETING ANALYTICS (Protiviti-Oriented)

#	**Question / KPI**	**Sample SQL Query / Explanation**
1	Calculate Click-Through Rate (CTR).	SELECT CampaignID, SUM(Clicks)*1.0/SUM(Impressions) AS CTR FROM Campaign GROUP BY CampaignID;

2	Find cost per lead.	SELECT CampaignID, SUM(Spend)/SUM(Leads) AS CostPerLead FROM Campaign GROUP BY CampaignID;
3	Identify top-performing campaigns by ROI.	SELECT CampaignID, (SUM(Revenue)-SUM(Spend))/SUM(Spend) AS ROI FROM Campaign GROUP BY CampaignID ORDER BY ROI DESC;
4	Calculate customer retention rate.	SELECT COUNT(DISTINCT CurrentYear.CustomerID)*100.0/COUNT(DISTINCT PrevYear.CustomerID) AS RetentionRate FROM Sales CurrentYear JOIN Sales PrevYear ON CurrentYear.CustomerID=PrevYear.CustomerID;
5	Find total conversions per channel.	SELECT Channel, SUM(Conversions) AS TotalConversions FROM Campaign GROUP BY Channel;
6	Compare monthly ad spend.	SELECT MONTH(Date) AS Month, SUM(Spend) AS TotalSpend FROM Campaign GROUP BY MONTH(Date) ORDER BY Month;
7	Show customers with multiple campaign responses.	SELECT CustomerID, COUNT(DISTINCT CampaignID) AS Campaigns FROM Responses GROUP BY CustomerID HAVING COUNT(DISTINCT CampaignID)>1;
8	Calculate average order value.	SELECT CustomerID, AVG(OrderValue) AS AOV FROM Orders GROUP BY CustomerID;
9	Identify churned customers (no purchase in last 6 months).	SELECT CustomerID FROM Customers WHERE CustomerID NOT IN (SELECT DISTINCT CustomerID FROM Orders WHERE OrderDate >= DATEADD(MONTH, -6, GETDATE()));

10	Rank regions by marketing ROI.	SELECT Region, ROI = (SUM(Revenue)-SUM(Spend))/SUM(Spend), RANK() OVER(ORDER BY (SUM(Revenue)-SUM(Spend))/SUM(Spend) DESC) AS ROI_Rank FROM Campaign GROUP BY Region;

V. FINANCE & MARKETING KPI SUMMARY (for SQL-based Questions)

Domain	**KPI**	**SQL Expression / Metric**
Finance	ROI	(SUM(Revenue)-SUM(Investment))/SUM(Investment)
	Profit Margin	(SUM(Revenue)-SUM(Cost))/SUM(Revenue)
	Net Profit	SUM(Revenue)-SUM(Expense)
	Working Capital	SUM(CurrentAssets)-SUM(CurrentLiabilities)
	Debt-to-Equity	SUM(TotalDebt)/SUM(Equity)
Marketing	CTR	SUM(Clicks)/SUM(Impressions)
	Conversion Rate	SUM(Conversions)/SUM(Leads)
	CAC	SUM(Spend)/COUNT(DISTINCT CustomerID)
	CLV	SUM(RevenuePerCustomer) or via subquery
	ROI per Campaign	(SUM(Revenue)-SUM(Spend))/SUM(Spend)

VI. SAMPLE PROTIVITI INTERVIEW QUESTIONS (SQL)

Question	**What They Assess**
How would you find the top 5 performing products by sales?	Sorting, grouping, and aggregation logic
Write a query to find customers who purchased in both 2023 and 2024.	JOINs and DISTINCT understanding
Explain the difference between INNER JOIN and LEFT JOIN with an example.	Relational model clarity

How would you calculate marketing ROI using SQL?	KPI application in business context
What is the use of window functions in business analytics?	Analytical SQL knowledge
How do you handle NULL values in financial data?	Data cleaning and preprocessing
Write a query to calculate month-over-month revenue growth.	Time-based aggregation
What is the difference between correlated and non-correlated subqueries?	Query optimization and logic understanding
How can SQL be integrated with Power BI?	Practical tool integration (Get Data → SQL Server connector)
Which functions are used for date/time analytics?	MONTH(), YEAR(), DATEADD(), DATEDIFF()

VII. QUICK REVISION CHEAT-SHEET

Topic	**Command / Example**
Select all columns	SELECT * FROM Table;
Filter records	SELECT * FROM Table WHERE Condition;
Aggregate	SELECT Column, SUM(Value) FROM Table GROUP BY Column;
Join two tables	SELECT * FROM A JOIN B ON A.ID=B.ID;
Subquery	SELECT * FROM Table WHERE Col>(SELECT AVG(Col) FROM Table);
Rank	RANK() OVER(ORDER BY Value DESC)
Date filter	WHERE YEAR(Date)=2024
Limit results	SELECT TOP 10 * FROM Table;

Handle missing	COALESCE(Column,0)
Create View	CREATE VIEW ViewName AS SELECT ...