📊 Data Modeling and Visualization (DMV) Viva Questions & Answers

This document provides short, high-probability oral questions and concise answers for each lab assignment, designed for quick review.

💾 Assignment 1: Data Loading, Cleaning & Transformation

Manual Questions

- **What is Data Analysis?**
  - The process of cleaning, changing, and processing raw data to extract **actionable, relevant information** for informed decision-making.
- **What is Data Cleaning?**
  - The process of identifying and fixing incorrect data (e.g., duplicates, corrupt, incomplete) to ensure accuracy and reliability.
- **What is Data Transformation?**
  - The process of converting **raw data** into a format or structure that is more suitable for model building, data discovery, and analysis.

Related Questions

- **Name two common steps in data cleaning.**
  - 1. Removing duplicates. 2. Handling missing values.
- **What is a unified format?**
  - Converting data loaded from various sources (CSV, Excel, JSON) into a **single, common data structure** (like a Pandas DataFrame) for seamless analysis.

☁️ Assignment 2: Interacting with Web APIs

Manual Questions

- **What is OpenWeatherMap?**
  - A **popular online service** that provides weather data through its API (Application Programming Interface).
- **What is a Web API?**
  - An **Application Programming Interface** (API) that provides a **standardized way** for different software systems to communicate and exchange data over the internet.
- **How do we interact with Web APIs?**

- By using an **HTTP Client Library** (e.g., Python `requests`) to make **API Requests** (GET/POST) after obtaining an **API Key** (if required) and parsing the JSON/XML response.

Related Questions

- **Why are API keys used?**
  - To **authenticate API requests** and allow the service provider to track usage against predefined limits.
- **What is Rate Limiting?**
  - A mechanism used by APIs to restrict the number of requests an application can make within a certain time frame to prevent abuse and ensure fair usage.

👥 Assignment 3: Data Cleaning & Preparation (Customer Churn)

Manual Questions

- **What is Data Preparation?**
  - The process of collecting, cleaning, labeling, and transforming raw data to make it suitable for further processing and machine learning analysis.
- **What is Customer Churn?**
  - The number or percentage of customers who **stop using** a company's products or services during a specific period.

Related Questions

- **Why is predicting churn important?**
  - It allows the company to **proactively identify customers at high risk** of leaving and implement targeted retention strategies (like personalized offers) to reduce customer loss.
- **Name two strategies for handling missing values.**
  - **Imputation** (filling missing data with mean, median, or mode) or **Removal** (deleting the rows/columns with missing data).

🏠 Assignment 4: Data Wrangling (Real Estate)

Manual Questions

- **What is Data Wrangling?**
  - The process of gathering, collecting, and **transforming raw data** into a clean, organized, and usable format for analysis and decision-making.
- **Name three steps in data wrangling.**
  - **Data Structuring**, **Data Cleaning**, and **Data Validating** (or Data Discovery, Enriching, Publishing).

Related Questions

- **What is the purpose of Feature Engineering in real estate data?**
  - To create new, informative variables (features) from existing data, such as calculating **price per square foot** or property age.
- **What is the goal of handling categorical variables?**
  - To convert non-numeric categories (like 'Property Type') into numerical representations using techniques like **One-Hot Encoding** or **Label Encoding** so they can be used in models.

📊 Assignment 5: Data Visualization (Matplotlib)

Manual Questions

- **What do you mean by data visualization?**
  - The **graphical representation** of information and data to help users understand **patterns, trends, and outliers** hidden within the data.
- **What is Matplotlib?**
  - A **comprehensive data visualization library in Python** used to create static, interactive, and animated plots.
- **What are Figure and Axes in Matplotlib?**
  - **Figure** is the **top-level container** (the canvas), and **Axes** is the **actual plotting area** where data is drawn (the subplot).

Related Questions

- **Name three types of plots for analyzing AQI Trends over time.**
  - **Line plots**, **Time series plots**, and **Bar plots**.

📈 Assignment 6: Data Aggregation (Retail Sales)

Manual Questions

- **What is Data Aggregation?**
  - The process of **collecting data** (from multiple sources) to present it in a **summary form** (e.g., total sales, average profit) for statistical analysis.
- **What are the different steps performed in data aggregation?**
  - **Gather Data**, **Define Regions**, **Data Aggregation** (summing, averaging), **Visualize the Data**, and **Compare Regions**.

Related Questions

- **What is the main objective of analyzing sales performance by region?**

    - To **identify high-performing and underperforming regions** to effectively allocate resources and target marketing/sales strategies.
- **Name two types of data aggregation.**
    - **Manual** Data Aggregation and **Automated** Data Aggregation.

🕰️ Assignment 7: Time Series Data Analysis (Stock Market)

Manual Questions

- **What is Time Series Analysis?**
    - A statistical technique that deals with **time series data** (a series of data points ordered in time) to identify trends, patterns, and potential predictors.
- **How is Analysis and Visualization of Stock Market Data performed?**
    - By plotting **time series plots** (line charts), calculating **moving averages** to smooth noise, and using models like **ARIMA** or **exponential smoothing** for forecasting.
- **Why is data visualization important?**
    - It helps people **see, interact with, and better understand data**, enabling effective communication of insights and **data-driven decision-making**.

Related Questions

- **What is the purpose of a moving average?**
    - To **identify the underlying trends** and **smooth out noise** (short-term fluctuations) in the time series data.