Assignment

Q1 Define Data Analytics. Discuss its importance in bussiness decision-making. Mention at least two real-world appl?

Soln Data analytics refers to the process of collecting, organizing, analyzing and interpreting data to uncover patterns, trends, and insights. It helps businesses make informed decisions by turning raw data into actionable knowledge. In today's data-driven world, companies use analytics to improve performance, understand customer behaviour, optimize operations and gain a competetive edge. It supports strategic planning and reduces risks by backing decisions with evidence.

Real-world applications-

(i) Retail - Companies like Amazon use data analytics to recommend products based on customer browsing & purchase history, increasing sales.

(ii) Healthcare - Hospitals use analytics to predict patient admission rates and optimize resource allocation, improving efficiency & outcomes.

Q2 Write SQL queries for the following tasks using a sample Sales table.

(a) Retrieve total sales per region.

Query:-
```
SELECT region, SUM(sales_amount) AS total_sales
FROM sales
GROUP BY region;
```

(b) Find the top 3 products with the highest sales.

Query:- SELECT product, SUM(sales_amount) AS total_sales
FROM sales
GROUP BY product
ORDER BY total_sales DESC
LIMIT 3;

(c) Display customer names who made purchases in the last 30 days.

Query:- SELECT DISTINCT customer_name
FROM sales
WHERE purchase_date >= CURRENT_DATE - INTERVAL 30 DAY;

Q3 What are the key steps involved in data cleaning and preparation? Explain each step with an example.

Sol'n Key steps involved in data cleaning and preparation are:

(i) Removing Duplicates - Identical records can distort analysis. for instance, if a customer order appears twice, it inflates sales fig. Tools like SQL's DISTINCT help fix this.

(ii) Handling Missing Data - Missing values can be filled using mean, median or predictive models. eg. if age data is missing, fill it with average age or estimate using similar profiles.

(iii) Correcting Data types - Ensure numerical data isn't stored as text. In SQL, cast data types appropriately.

(iv) Standardizing formats - Data like dates may appear inconsistently, standardizing improves processing.

(v) Outlier Detection - Identify anomalies, such as transaction with negative value, using box plots or standard deviation.

Q4 You are given an Excel dataset with missing values, duplicate rows and inconsistent date formats. Describe how you would clean and prepare this data using Advanced Excel techniques.

Soln 4 To clean an Excel dataset with missing values, duplicates and inconsistent date formats, following these steps:-

(i) Remove duplicates- Use the "Remove Duplicates" tool under the Data tab. Select relevant columns to eliminate repeated rows without affecting unique data.

(ii) Handle Missing Values- Use formula like =IF(ISBLANK(A2), AVERAGE(A:A), A2) to fill gaps with averages. You can also use Flash fill or Power query to infer patterns and auto-complete data.

(iii) Standardize Data formats- Apply consistent date formats using the format cells dialog or use TEXT(A2, "DD-MM-YYYY") for uniformity. Power Query is also effective for parsing inconsistent dates.

(iv) Validation checks- Use conditional formatting to highlight blanks or anomalies. Create data ~~validation~~ validation rules to prevent future entry errors.

Q5 List and Explain at least four common data visualization techniques. Specify when to use each type of chart (bar chart, line chart, pie chart, scatter plot).

Soln 5 (i) Bar chart- Best for comparing categories such as sales across regions or products types. It's ideal when dealing with discrete data.

(ii) Line chart- Used for showing trends over time. eg. monthly sales or website traffic patterns. Suitable for continous data.

(iii) Pie chart- shows part of whole, such as market share distribution. Use it when you have a small number of categories & want to display proportions. Avoid it for complex datasets.

(iv) Scatter Plot- Useful for identifying relationships b/w two variables, such as advertising spend vs revenue. Helpful in regression analysis and outlier detection.

Q6 Explain how Pivot Tables and VLOOKUP/XLOOKUP in Excel can be used for data analysis. Provide a use case scenario for each.

Sol^n6 Pivot tables - Allow users to summarize, group and analy large datasets quickly. eg. you can create a pivot tab to show total sales by product & region, filter by dat or calculate averages. It transform raw data into insightful summaries without complex formulas.

VLOOKUP/XLOOKUP - Used to retrieve related information from another table. For instance, VLOOKUP can match a Product ID in one table to its price in another. XLOOKUP is more flexible and supports searches in any direction handling missing values more gracefully.

Use case:

- Pivot table - Analyzing employee hours per departmat
- VLOOKUP/XLOOKUP - Pulling employee names based on ID from a master list.

Q7 Compare Power BI and Excel for data visualization. In what situations would one be preferred over the other?

Sol^n7 Power BI is a powerful business intelligence tool designed for interactive, real-time dashboards and handling large datasets. It supports advanced visualizations, DAX formulas, and easy data integration from multiple sources. It's ideal for sharing reports across teams & automating updates.

Excel excels at manual analysis, quick data manipulation and ad-hoc reporting. It's widely used for detailed data inspection & simpler visualizations.

When to use -

- Power BI - for dynamic dashboards, enterprise-level reporting and connecting to databases or cloud data sources.

- Excel - for data entry, one-off analyses or when the dataset is small or local.

Q8 Explain the importance of data types in SQL and Excel during data analysis. How do incorrect data types affects results & visualizations?

Sol^n 8 In SQL - Using wrong data types can cause errors or incorrect results. eg. storing dates as strings prevents proper date comparisons or aggregations. Numeric calculations fail if values are saved as text.

In Excel - Incorrect types can break formulas or mislead charts. for instance, if dates are stored as text, functions like MONTH() or DATEDIF() won't work. sorting or filtering may also give unexpected results.

Impact on Analysis:-

- Misleading Results.
- Broken formulas or queries.
- Inaccurate visualizations.

Completed