

Cloudera

CDP-4001

CDP Data Analyst- Certification Exam

For More Information – Visit link below:

<https://www.examsempire.com/>

Product Version

1. Up to Date products, reliable and verified.
2. Questions and Answers in PDF Format.



<https://examsempire.com/>

Visit us at: <https://www.examsempire.com/cdp-4001>

Latest Version: 6.1

Question: 1

You are analyzing a dataset with sales data for different product categories over several years. You want to visualize the trend of sales for each category over time, highlighting any significant changes or anomalies. Which visualization type would be most effective for this purpose?

- A. Scatter Plot
- B. Line Chart
- C. Bar Chart
- D. Pie Chart
- E. Heatmap

Answer: B

Explanation:

A line chart is the most appropriate visualization for showing trends over time. It effectively depicts the changes in sales values for each product category over the years, making it easy to identify any significant increases, decreases, or anomalies.

Question: 2

You are investigating the distribution of customer ages in a retail dataset. You want to identify the age groups with the highest customer concentration and visualize the distribution in a way that emphasizes the density of each age range. Which visualization would be most suitable for this purpose?

- A. Histogram
- B. Scatter Plot
- C. Bar Chart
- D. Pie Chart
- E. Box Plot

Answer: A

Explanation:

A histogram is the best choice for visualizing the distribution of numerical data, like customer ages. It groups the data into bins and displays the frequency of values within each bin, highlighting the density of age ranges and identifying the age groups with the highest customer concentration.

Question: 3

You are analyzing a dataset containing information about website user behavior, including page views, click-through rates, and time spent on each page. You want to create a visualization that shows the relationships between these metrics and helps identify any patterns or correlations. Which visualization technique would be most appropriate?

- A. Line Chart
- B. Scatter Plot
- C. Heatmap
- D. Pie Chart
- E. Parallel Coordinates Plot

Answer: E

Explanation:

A parallel coordinates plot is well-suited for visualizing relationships between multiple variables, like page views, click-through rates, and time spent on a page. It displays each variable on a separate axis, allowing you to see how values change across different dimensions, making it easier to identify patterns and correlations.

Question: 4

You are analyzing a dataset with customer feedback ratings for different products. You want to create a visualization that allows users to easily filter and explore the data based on product category, rating, and other relevant criteria

a. Which visualization type would be most suitable for this interactive exploration?

- A. Bar Chart
- B. Line Chart
- C. Scatter Plot
- D. Interactive Dashboard
- E. Treemap

Answer: D

Explanation:

An interactive dashboard is the most suitable option for exploring customer feedback data. It allows users to dynamically filter and slice the data based on product category, rating, and other criteria, providing a flexible and intuitive way to analyze the feedback and gain insights.

Question: 5

You are working on a visualization to demonstrate the distribution of customer purchase frequencies. You want to show the number of customers who purchase products once, twice, three times, and so on, within a specific time period. Which visualization would be most suitable for this purpose?

- A. Scatter Plot
- B. Line Chart

- C. Bar Chart
- D. Pie Chart
- E. Histogram

Answer: C

Explanation:

A bar chart is the most effective visualization for displaying the distribution of customer purchase frequencies. It allows you to represent the number of customers in each frequency category (once, twice, etc.) as bars, clearly showing the concentration of customers at different purchase frequency levels.

Question: 6

You are building a dashboard to track website traffic patterns. You want to display a chart that shows the average number of visitors per hour over the past 24 hours, grouped by the day of the week. Which Cloudera Data Visualization component would be the most appropriate for this task?

- A. Bar Chart
- B. Line Chart
- C. Scatter Plot
- D. Pie Chart
- E. Heatmap

Answer: B

Explanation:

A line chart is the most suitable choice for displaying trends over time, especially when grouped by categories like days of the week. It clearly shows the average visitor count per hour, allowing for easy identification of patterns and variations throughout the day.

Question: 7

You need to create a dashboard that presents a summary of key performance indicators (KPIs) for a large e-commerce platform. The KPIs include daily sales revenue, average order value, and the number of new customers. Which Cloudera Data Visualization component would be most effective for displaying these KPIs in a visually appealing and informative way?

- A. Gauge Chart
- B. Table
- C. Bullet Chart
- D. Treemap
- E. Stacked Bar Chart

Answer: A,C

Explanation:

Gauge charts are ideal for displaying single KPIs with a clear target or benchmark, providing a visual representation of progress towards that goal. Bullet charts, on the other hand, are excellent for comparing a primary measure (e.g., sales revenue) against a benchmark or target, while also displaying additional context like the range of possible values. Both components offer a visually engaging and insightful way to present the KPIs.

Question: 8

You're analyzing a dataset of website activity in Impala, and need to identify the tables that store information about specific page views.

You are aware that the dataset has different naming conventions for tables related to page views, such as 'pageviews', 'page_hits', and 'website activity'. How would you effectively list all tables relevant to your analysis?

- SHOW TABLES LIKE 'page%';
- SHOW TABLES LIKE '%page%';
- SHOW TABLES LIKE '%page%' OR '%hits%' OR '%activity%';
- DESCRIBE DATABASE ;
- SELECT * FROM INFORMATION_SCHEMA.TABLES WHERE TABLE_NAME LIKE '%page%';

- A. Option A
- B. Option B
- C. Option C
- D. Option D
- E. Option E

Answer: C,E

Explanation:

Options C and E are the most effective for this scenario. Option C utilizes wildcards to match table names containing 'page', 'hits', or 'activity', capturing different naming conventions. Option E queries the INFORMATION_SCHEMA.TABLES to filter tables based on the '%page%' pattern. Option A would only find tables starting with 'page', missing other potential tables. Option B would return tables ending with 'page', not matching tables like 'pageviews'. Option D would list all tables in the database, not specifically focusing on page view data.

Question: 9

You have a Hive table with a column 'product id' that contains both numeric and alphanumeric values. You need to extract the numeric part from the column, converting it to an integer Which HiveQL statement achieves this?

- SELECT CAST(product_id AS INT) FROM my_table;
- SELECT CAST(REGEXP_REPLACE(product_id, '[^0-9]', '') AS INT) FROM my_table;
- SELECT CAST(SUBSTR(product_id, 1, LENGTH(product_id) - 1) AS INT) FROM my_table;
- SELECT CAST(product_id AS STRING) FROM my_table WHERE LENGTH(product_id) > 0;
- SELECT CAST(product_id AS INT) FROM my_table WHERE REGEXP_LIKE(product_id, '[0-9]+');

- A. Option A
- B. Option B
- C. Option C
- D. Option D
- E. Option E

Answer: B

Explanation:

The correct statement is option B: "SELECT CAST(REGEXP_REPLACE(product_id, '[^0-9]', '') AS INT) FROM my_table;". This statement uses REGEXP_REPLACE to remove all non-numeric characters and then casts the result to an INT. The other options are incorrect because: * **Option A:** "SELECT CAST(product_id AS INT) FROM my_table;" tries to cast the entire string, which will fail if the column contains non-numeric values. * **Option C:** "SELECT CAST(SUBSTR(product_id, 1, LENGTH(product_id) - 1) AS INT) FROM my_table;" extracts a substring but doesn't handle the conversion to an integer. * **Option D:** "SELECT CAST(product_id AS STRING) FROM my_table WHERE LENGTH(product_id) > 0;" simply casts the column to a STRING without extracting numeric values. * **Option E:** "SELECT CAST(product_id AS INT) FROM my_table WHERE REGEXP_LIKE(product_id, '[0-9]+');" filters out rows with non-numeric values but doesn't extract the numeric portion.

Thank You for Trying Our Product
Special 16 USD Discount Coupon: NSZUBG3X
Email: support@examsempire.com

**Check our Customer Testimonials and ratings
available on every product page.**

Visit our website.

<https://examsempire.com/>