2024

BRICS SKILLS COMPETITION

(BRICS FUTURE SKILLS & TECHNOLOGY CHALLENGE)

Data Analysis and Visualization

BRICS-FS-36

Test Project

(International Final)

July 2024

Contents

1. Form of participation

The competition format for this event is individual competition.

2. Content of the competition

This competition consists of 4 modules, and participants are required to complete all competition tasks in order. During the competition, unified competition topic files, competition equipment, basic operation instruction files for the equipment, as well as data sources or other technical prerequisites needed to ensure the independence and fairness of each task module, will be provided to the participants.

The competition content includes the following task modules based on data analysis and visualization:

Module A: Data Acquisition and Processing

Module B: Data Analysis and Operation

Module C: Data Presentation and Sharing

Module D: Data Development and Application

If participants fail to comply with occupational health and safety requirements or pose risks to themselves and other participants, they may be disqualified from the competition.

After completing the competition, the referee team will score the results submitted by the participants.

3. Project modules and time requirements

3.1 Project modules and time requirements

The data analysis and visualization competition consists of 4 modules, with a total time limit of 480 minutes for participants. Specific project module names and time requirements can be found in Table 1: Project Modules and Time Requirements Checklist

Table 1: Project Modules and Time Requirements Checklist

Serial number	Module name	Completion time for competition content
1	Module A: Data Acquisition and Processing	120min
2	Module B: Data Analysis and Operation	120min
3	Module C: Data Presentation and Sharing	120min
4	Module D: Data Development and Application	120min

3.2 Task content

Module A: Data Acquisition and Processing (120 minutes)

Module Description

As scientific and technological advancements continue to evolve and update, people can more easily access the information they need from various sources. With the proliferation of information in people's lives, the era of big data has emerged. In the context of big data, the quantity and complexity of medical data have also been continuously increasing. In recent years, with the popularization of electronic medical records and the advancement of information technology, medical data analysis has become increasingly important. Medical data analysis has a wide range of applications in medical research, clinical decision support, disease prevention, and public health.

For example, medical data analysis can be used to study the pathogenesis of diseases, explore new treatment methods, evaluate healthcare policies, and plan public health strategies. At the same time, medical data analysis can also be used to improve the efficiency and quality of healthcare, such as optimizing the allocation of medical resources, reducing medical errors, and increasing patient satisfaction.

Big data is of great significance to healthcare. Healthcare institutions can effectively help doctors make more accurate clinical diagnoses, predict the cost and effectiveness of treatment plans more accurately, integrate patient genetic information for personalized treatment, analyze population health data to predict disease outbreaks, and so on.

Task One: Data Processing

1. Operate in the "Medical Data" worksheet, delete duplicate rows of data, keeping the first occurrence when removing duplicates, and save the completed data after the operation.

2. Count the number of missing values in the range A1:N16551 of the "Medical Data" worksheet and save the answer to the corresponding position in the "answer" column of the "results" table.

3. Operate in the "Medical Data" worksheet, handle missing values according to the following requirements:

·Delete rows with missing values in the age or gender fields.

·Fill missing values in the regional division field with values from the treatment area.

4. Operate in the "Medical Data" worksheet, handle the data in the "diagnosis_name" field according to the standards specified in the "diagnosis_process" worksheet, and save the completed data after the operation.

5. Operate in the "Medical Data" worksheet, convert the timestamp in the "settlement_date" field into a date format (e.g., 20180101), and save the completed data after the operation.

6. Operate in the "Medical Data" worksheet, divide the data in the "age" field into levels according to the following criteria and create a new field called "age_level," then save the completed data after the operation:

Infants: $0 < age \leq 6$

Children: $6 < age \leq 12$

Adolescents: $12 < age \leq 17$

Youth: $17 < age \leq 45$

Middle-aged: $45 < age \leq 69$

Elderly: $69 < age < 100$

Task Two: Data Analysis

1. Count the age group with the highest number of patients in the "age_level" field of the "Medical Data" worksheet, and save the answer to the corresponding position in the "answer" column of the "results" table.

2. In the "Medical Data" worksheet, identify the medical type with the highest total cost, and save the answer to the corresponding position in the "answer" column of the "results" table.

3. Analyze the proportion of insured individuals across regions seeking medical treatment each year based on admission dates in the "Medical Data" worksheet (in decimal form, rounded to two decimal places), and save the answer to the corresponding position in the "2.3" table.

4. Calculate the proportion of retirees to the total number of individuals in the "Medical Data" worksheet (in percentage form, rounded to two decimal places), and save the answer to the corresponding position in the "answer" column of the "results" table.

5. Identify the person ID with the highest number of medical visits in the "Medical Data" worksheet. Save the person ID and the number of medical visits to the corresponding position in the "answer" column of the "results" table.

6. Calculate the length of hospital stay based on admission dates and discharge times in the "Medical Data" worksheet, and identify the disease with the longest average hospital stay. Save the answer to the corresponding position in the "answer" column of the "results" table.

7. Count the number of diseases with an average reimbursement rate exceeding 80% in the "Medical Data" worksheet. Save the answer to the corresponding position in the "answer" column of the "results" table.

Average reimbursement rate = Average basic medical insurance fund amount / Average total cost

Task Three: Data Visualization

1. Analyze the distribution of medical visits across different age groups based on data from the "Medical Data" worksheet.

·Present the analysis in the form of a pie chart.

·Display data labels with two decimal places.

·Save the chart to the "3.1" worksheet.

2. Analyze the total expenditure and total basic medical insurance fund amount for different personnel types based on data from the "Medical Data" worksheet.

·Present the analysis in the form of a bar chart and a line chart.

·Total expenditure is represented as a bar chart, while the total basic medical insurance fund amount is represented as a line chart.

·Arrange the data in descending order of total expenditure.

·Set the maximum value of the y-axis to 21,000,000 and the minimum value to 0.

·Save the chart to the "3.2" worksheet.

3. Analyze the average personal expenditure for different tumor diseases based on data from the "Medical Data" worksheet.

·Calculate personal expenditure as the total expenditure minus the basic medical insurance fund amount.

·Present the analysis in the form of a bubble chart.

·Represent bubble size with the average personal expenditure.

·Display labels for the top five diseases with the highest average personal expenditure, centered, and assign bubble colors in the order: FFCCCC, 99CCCC, CCCC99, CC99CC, CC99FF.

·Save the chart to the "3.3" worksheet.

Module B: Data Analysis and Operation (120 minutes)

Module Description:

Under the impact of big data technology and concepts, education is undergoing a "silent revolution," with the transformation of teaching paradigms serving as its vanguard and core. With the advent of the big data era, teaching paradigms have also entered the 3.0 era. Campus data analysis refers to the process of using various technologies and tools to deeply explore and analyze various data within schools, thereby providing scientific basis for school decision-making and management. In today's information age, school management faces various challenges and opportunities. For example, student enrollment, teaching quality assessment, teacher-student behavior supervision, financial management, etc., all require data support for decision-making. Actively exploring education big data-driven teaching paradigms empowers teachers with microscopic observational abilities and telescopic predictive capabilities using real teaching data. This enables teaching to become scientific, intelligent, precise, and personalized.

Task One: Data Processing

1.1 Read the data, read the data from the four tables `Consumption.csv`, `attendance.csv`, `teacher.csv`, and `mark.csv`, and save them to variables `data_Consumption`, `data_attendance`, `data_teacher`, and `data_mark` respectively. Run the provided answer-saving code to save the answers.

1.2 Analyze the variable `data_Consumption`, process the `DealTime` column into date format, update the processed result to the variable `data_Consumption`, and run the provided answer-saving code to save the answers.

1.3 Analyze the variable `data_Consumption`, convert the `MonDeal` column to positive numbers, update the processed result to the variable `data_Consumption`, and run the provided answer-saving code to save the answers.

1.4 Analyze the variable `data_attendance`, process the data for February 2014, update the `qj_term` to '201320142', update the processed result to the variable `data_attendance`, and run the provided answer-saving code to save the answers.

1.5 Analyze the columns `DataDateTime` and `qj_term` in the variable `data_attendance`, find the relationship between `DataDateTime` and `qj_term`, fill in the missing term information, return the term with the most attendance and the corresponding attendance count, save the result to `task1_5`, and run the provided answer-saving code to save the answers.

Task Two: Campus Teacher Information Analysis

2.1 Analyze the columns `sub_Name` and `bas_id` in the variable `data_teacher`, count the number of teachers in each subject, return the subject with the most teachers, save the result to the variable `task2_1`, and run the provided answer-saving code to save the answers.

2.2 Analyze the columns `term`, `gra_Name`, and `bas_id` in the variable `data_teacher`, observe the teaching situation of teachers in the "201420151" semester, count how many teachers are involved in teaching multiple grades, save the result to the variable `task2_2`, and run the provided answer-saving code to save the answers.

2.3 Analyze the columns `term`, `cla_Name`, and `bas_id` in the variable `data_teacher`, observe the teaching situation of teachers in the "201420151" and

"201420152" semesters, count the number of classes taught by each teacher, return the `bas_id` of the teacher with the most classes, save the result to the variable `task2_3`, and run the provided answer-saving code to save the answers.

2.4 Analyze the columns `term`, `cla_Name`, and `gra_Name` in the variable `data_teacher`, count the number of classes for each semester and grade, return the semester and grade with the most classes, save the result to the variable `task2_4`, and run the provided answer-saving code to save the answers.

2.5 Analyze the variable `data_teacher`, count the number of classes for high school physics and geography subjects in the "201420151" semester, save the result in the format (number of geography classes, number of physics classes), save the result to the variable `task2_5`, and run the provided answer-saving code to save the answers.

Task Three: Campus Attendance Statistics

3.1 Analyze the variable `data_attendance`, count the number of attendances containing "School uniform" in the `controler_name` column, return the semester with the most late arrivals and early departures, save the result to the variable `task3_1`, and run the provided answer-saving code to save the answers.

3.2 Analyze the variable `data_attendance`, count the student ID with the most late arrivals and early departures in the first year, save the result to the variable `task3_2`, and run the provided answer-saving code to save the answers.

·Late arrival: `controler_name` contains "Late arrival"

·Early departure: `controler_name` contains "Early departure"

3.3 Analyze the variable `data_attendance`, calculate the proportion of late arrivals or early departures in the attendance records for `bf_classid` with attendance counts

exceeding 200, return the maximum proportion, save the result to the variable `task3_3`, and run the provided answer-saving code to save the answers.

3.4 Analyze the variable `data_attendance`, count the number of late arrivals and early departures per hour, return the peak time for late arrivals and early departures (in 24-hour format), save the result to the variable `task3_4`, and run the provided answer-saving code to save the answers.

3.5 Analyze the variable `data_attendance`, count the number of students who have been both late and early to class on the same day, save the result to the variable `task3_5`, and run the provided answer-saving code to save the answers.

Task Four: Student Performance Analysis

4.1 Analyze the variable `data_mark`, calculate the average score of the chemistry subject for each class of high school seniors in the exam with `exam_number` 289, return the `cla_id` with the highest average score, save the result to the variable `task4_1`, and run the code to save the answers.

4.2 Analyze the variable `data_mark`, divide the scores of the chemistry subject in the exam with `exam_number` 289 according to the following criteria, count the number of students in each score range, return the score range with the most students, save the result to the variable `task4_2`, and run the code to save the answers.

·Score = 3: Exempt

·2 <= Score < 60: Fail

·60 <= Score < 70: Pass

·70 <= Score < 80: Average

·80 <= Score < 90: Good

·90 <= Score <= 100: Excellent

4.3 Analyze the variable `data_mark`, calculate the Z-score for the group of data from the exam with `exam_number` 289 for the chemistry subject. Determine the maximum Z-score. Save the result to the variable `task4_3` and run the code to save the answers.

·The Z-score is a dimensionless quantity used to measure the difference between a raw score and the mean of a dataset. It is obtained by dividing the difference between the raw score and the dataset's mean by the standard deviation. The Z-score indicates the extent to which a raw score deviates from the mean, measured in terms of standard deviations. A positive Z-score indicates a score above the mean, while a negative Z-score indicates a score below the mean.

Z_score = (score - mean) / standard deviation

the result rounded to two decimal places.

4.4 Analyze the variable `data_mark`, count the number of cheating and absence incidents for each class in the exam with `exam_number` 284. Return the `cla_id` with the highest combined count of cheating and absence incidents. Save the result to the variable `task4_4` and run the code to save the answers.

4.5 Analyze the variable `data_mark`, calculate the number of times each class from high school juniors and seniors ranks first in each subject in the exam with `exam_number` 289. Return the `cla_id` with the highest frequency of ranking first. Save the result to the variable `task4_5` and run the code to save the answers.

4.6 Analyze the variable `data_mark`, calculate the total score of each student in the exam with `exam_number` 289 (rounding the score to one decimal place). Return the

highest total score among all grades. Save the result to the variable `task4_6` and run the code to save the answers.

·Total score `score` is composed of three parts:

·`score1`: Scores of Chinese, mathematics, and English

·`score2`: The highest score between physics and geography

·`score3`: The highest two scores among biology, politics, chemistry, and history

·score = score1+score2+score3

Module C: Data Presentation and Sharing (120 minutes)

Module Description:

With the development of enterprises and the competition in the market, the analysis of sales data has become increasingly important. Sales data analysis refers to the collection, organization, analysis, and utilization of sales performance data of enterprises to understand the company's sales performance, discover potential market opportunities, optimize sales strategies and marketing plans, in order to increase the company's sales revenue and profit.

In modern enterprises, sales data analysis has become an indispensable tool. Through data analysis, enterprises can better understand their market, customers, and competitors, thereby making wiser business decisions. These decisions may include pricing strategies, product promotion, improving customer service, and more. The RFM model is an important tool and method for measuring customer value and their ability to create profits. Among the numerous customer relationship management analysis models, the RFM model is widely mentioned. This model observes the customer's value status through three indicators: the customer's recent purchasing behavior, overall purchase frequency, and the amount spent.

Task One: Product Analysis

Task 1.1: Display the sales amount for products containing "diapers" in the product description.

Task Requirements:

Perform the following operations in the worksheet named "1.1" and save the final result:

Graph Name: Bar Chart

·The keyword "diapers" in the product description, consistent with the column labels in the example graph.

·The "Product Description" field as the column and the sales amount as the row. The column name of the product description should be exactly the same as in the example graph.

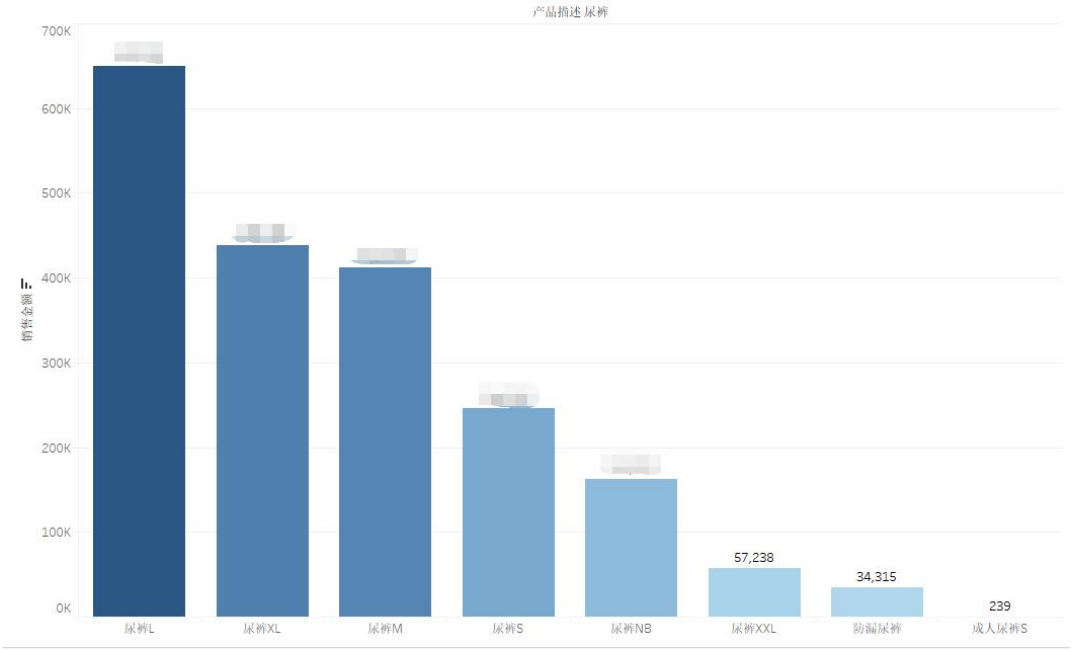
·Sort by sales amount from left to right in descending order.

·Display the sales amount as data labels.

·Set the color of the bars to blue, with darker shades indicating larger sales amounts, consistent with the example graph.

·Set the view size to 'Entire View'.

Reference Graph:



icon 1

Task 1.2: Display Sales Amount by Subcategory

Task Requirements:

Perform the following tasks on the worksheet named "1.2" and save the final result:

Chart Name: Word Cloud

Display labels for each subcategory, with the size of the label controlled by the sales amount.

Set the color of sales amount to tableau classic 20, consistent with the example.

Sort the subcategories by sales amount in ascending order.

Set the view size to "Entire View".

Reference image is as follows:

icon 2

Task 1.3: Display Sales Amount for Top 10 Categories

Task Requirements:

Perform the following tasks on the worksheet named "1.3" and save the final result:

Chart Name: Horizontal Bar

Sales amount field in columns, category field in rows

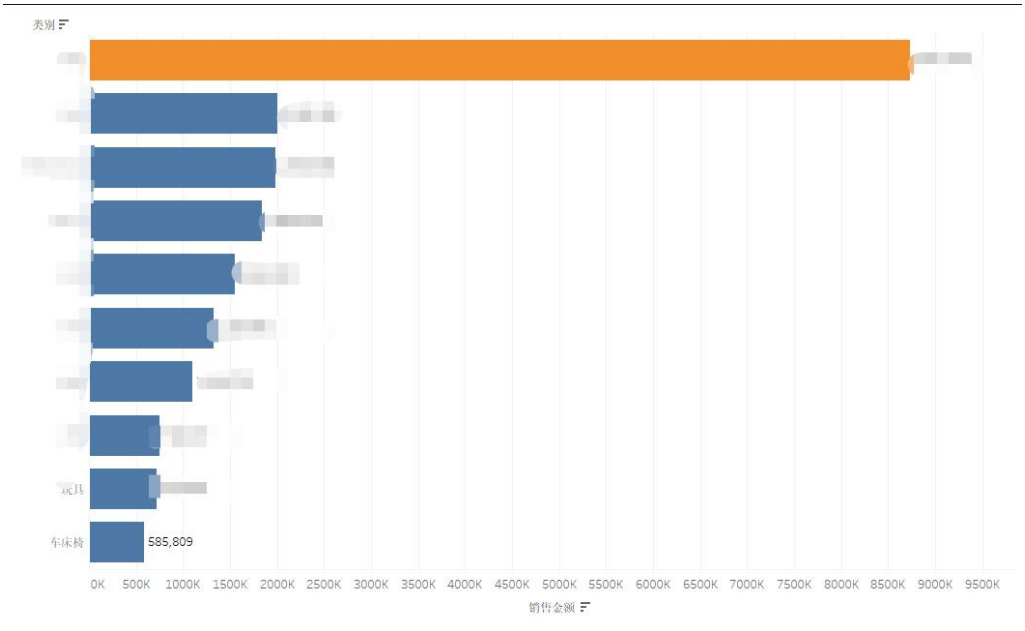
Only retain the top 10 categories by total sales amount. The category with the highest sales amount should be colored orange, while the remaining nine categories should be colored blue, consistent with the example.

Sort the categories by the total sum of sales amount in descending order.

Display sales amount labels as shown in the example.

Set the view size to "Entire View".

Reference image is as follows:



icon 3

Task 1.4: Display the Percentage of Food and Apparel Categories

Task Requirements:

Perform the following tasks on the worksheet named "1.4" and save the final result:

Chart Name: Donut Chart

For categories, apparel categories should include maternity wear, children's accessories, children's footwear, children's hosiery, children's shoes, children's clothing, and infant clothing; food categories include infant formula and food.

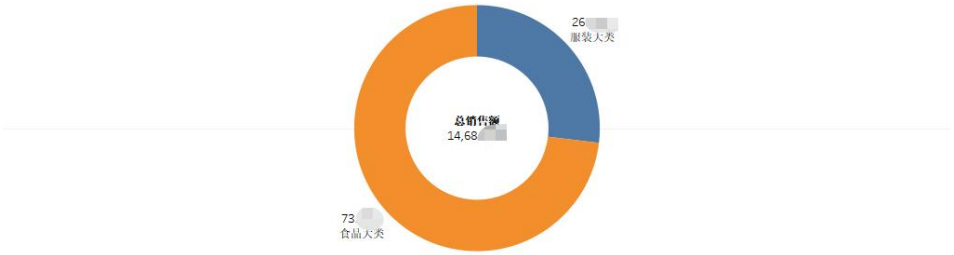
The angles of apparel and food categories represent different sales amounts.

Display labels for apparel, food categories, and sales amount percentages, retaining two decimal places, consistent with the example.

Set the inner label of the donut to "Total Sales Amount" to display labels for the total sales amounts of apparel and food categories, consistent with the example.

Apparel and food categories should be colored blue and orange, respectively, consistent with the example.

Reference image is as follows:



icon 4

Task 1.5: Display the Category Analysis Dashboard for 2013

Task Requirements:

Perform the following tasks on the dashboard named "2013 Category Analysis" and save the final result:

Dashboard Title: 2013 Category Analysis

Set the dashboard size to Desktop and apply the correct background image.

Stacked bar chart displays the top 6 categories by sales amount in 2013 and corresponding percentage labels, consistent with the example.

Tree map displays data for the year 2013, showing category and sales amount labels, sorted by category in ascending order of sales amount, consistent with the example.

Highlight table displays data for the year 2013, showing the top 10 categories and sales amount growth rate percentage data, sorted by category labels on the Y-axis, consistent with the example.

Color settings for the stacked bar chart, tree map, and highlight table should be red and gold, consistent with the example.

Set the title for the tree map and highlight table to the default font "tableau book", font size 15, and bold.

Reference image is as follows:

icon 5

Task Two: Order Analysis

Task 2.1: Display the sales distribution of order dates in 2012.

Task Requirements:

Perform the following tasks in the worksheet named "2.1" and save the final results:

Chart Name: Dual-Axis Chart

· Use the Order Date field for columns and the Sales Amount field for rows.

· Display only the category with the highest total sales amount.

· Show sales amount data for 2012, with January 2012 as the fixed date, comparing sales amounts from February to December 2012, and calculate the base growth rate.

Base Growth Rate = (Sales Amount for the Current Month / Sales Amount for the Fixed Month)

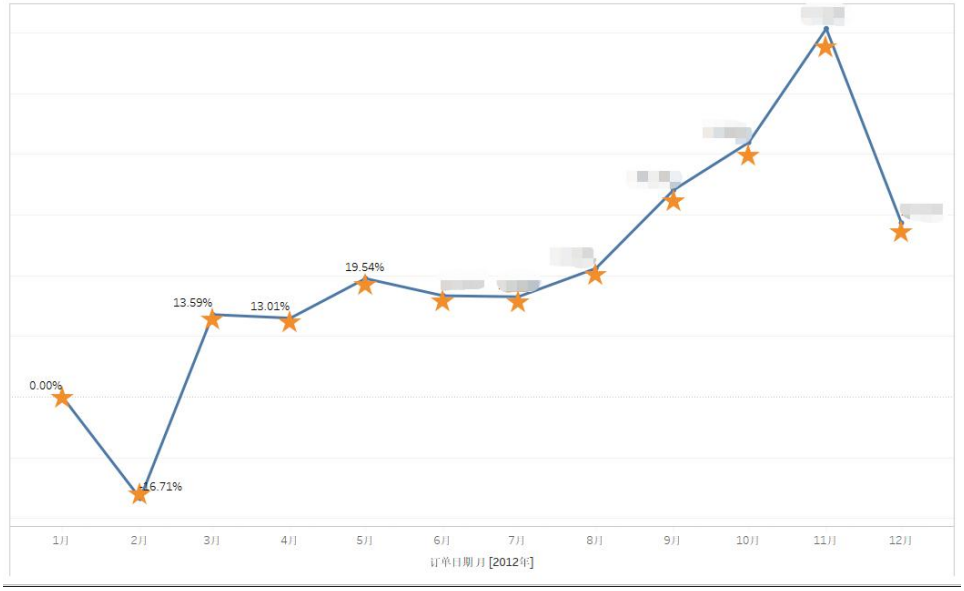
· Display the base growth rate as data labels, placing the labels above the line chart, consistent with the example.

· Do not display titles on both sides of the X-axis, consistent with the example.

· Set the line chart color to blue and the pentagram color to orange.

· Set the view size to "Entire View."

Reference Graph:



icon 6

Task 2.2: Use time series analysis to forecast sales amounts.

Task Requirements:

Perform the following tasks in the worksheet named "2.2" and save the final results:

Chart Name: Forecast Chart

· Use the Order Date field for columns and the Sales Amount field for rows, retaining only the category of toiletries.

· Keep the order dates from January 1, 2012, to July 24, 2014.

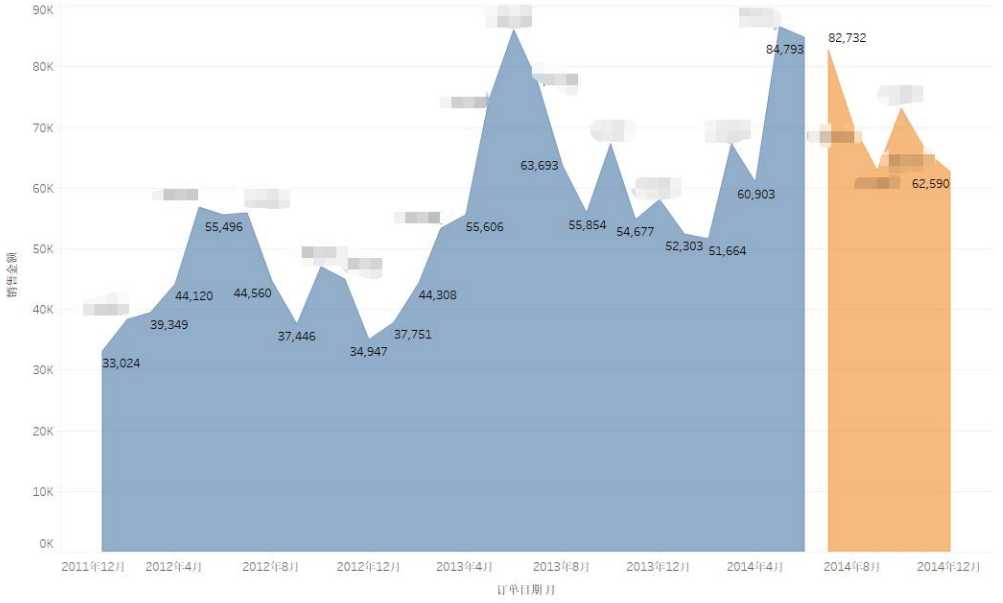
· Set the date format to months, forecast future 6 months of data, excluding the last month.

· Set the trend and seasonality to cumulative.

· Set the actual data to blue and the forecasted estimate data to orange, and set the view size to "Entire View."

· Display sales amount data labels, consistent with the example.

Reference Graph:



icon 7

Task Three: Comprehensive Analysis

Task 3.1: Display Membership Retention

Task Requirements:

Perform the following tasks in the designated worksheet and save the final results:

Chart Name: Funnel Chart

· Calculate the "Purchase Point Membership Lifespan (Months)" indicator, calculated in months. For example: Purchase Point Membership Lifespan = Membership Order Date (March 20, 2010) - Membership Creation Date (February 25, 2010) = 1.

· Use the count of User IDs for columns and the Purchase Point Membership Lifespan (Months) field for rows.

· Sort the Purchase Point Membership Lifespan (Months) in descending order based on the count of User IDs.

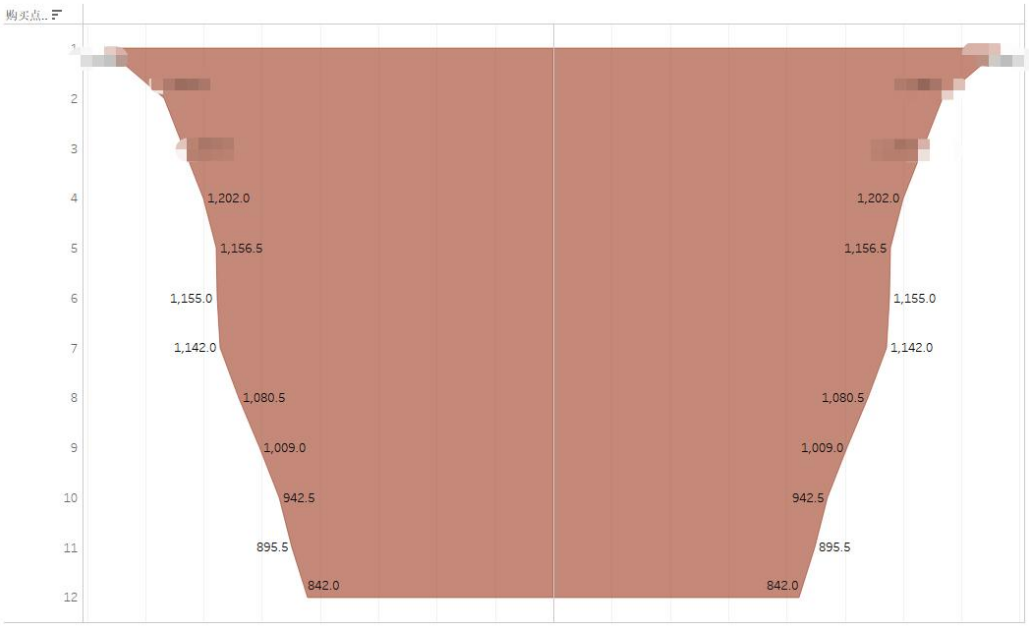
· Keep only the data for 1 month to 12 months of the Purchase Point Membership Lifespan.

· Display data labels for the count of User IDs as shown in the example.

· Set the color of the funnel chart to orange to match the example.

· Set the view size to "Entire View".

Reference Image:



icon 8

Task 3.2: Display the Number of Users Based on the Last Purchase Point Lifespan

Task Requirements:

Perform the following tasks in the designated worksheet and save the final results:

Chart Name: Histogram (Quality Distribution Chart)

· Calculate the "Purchase Point Membership Lifespan (Days)" indicator, where Purchase Point Membership Lifespan = Membership Order Date - Membership Creation Date (calculated in days).

· Calculate the "Last Purchase Point Lifespan (Days)" indicator. Based on this indicator, create a new field with bucket size of 30 days. Name this new field "Distribution of Last Purchase Point Lifespan".

· Use the "Distribution of Last Purchase Point Lifespan" field for columns and the User ID field for rows.

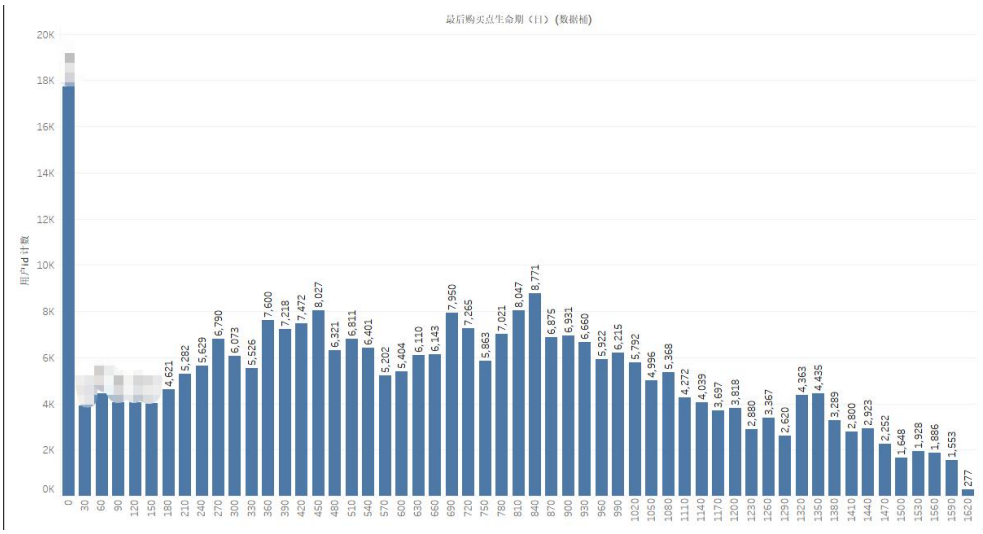
· Keep data for orders from July 1, 2010, to July 24, 2014.

· Display data labels for the count of User IDs.

· Set the color of the histogram to blue to match the example.

· Set the view size to "Entire View".

Reference Image:



icon 9

Task 3.3: Display User Purchase Frequency and Sales Amount

Task Requirements:

Perform the following tasks in the designated worksheet and save the final results:

Chart Name: Scatter Plot

· Calculate the "Total Orders Placed per Customer" indicator and create a new field. Name this new field "Total Orders Placed per Customer".

· Use the "Total Orders Placed per Customer" field for columns and the Sales Amount field for rows.

· Set a linear trend line to show the trend in sales amount based on the total orders placed per customer, consistent with the example.

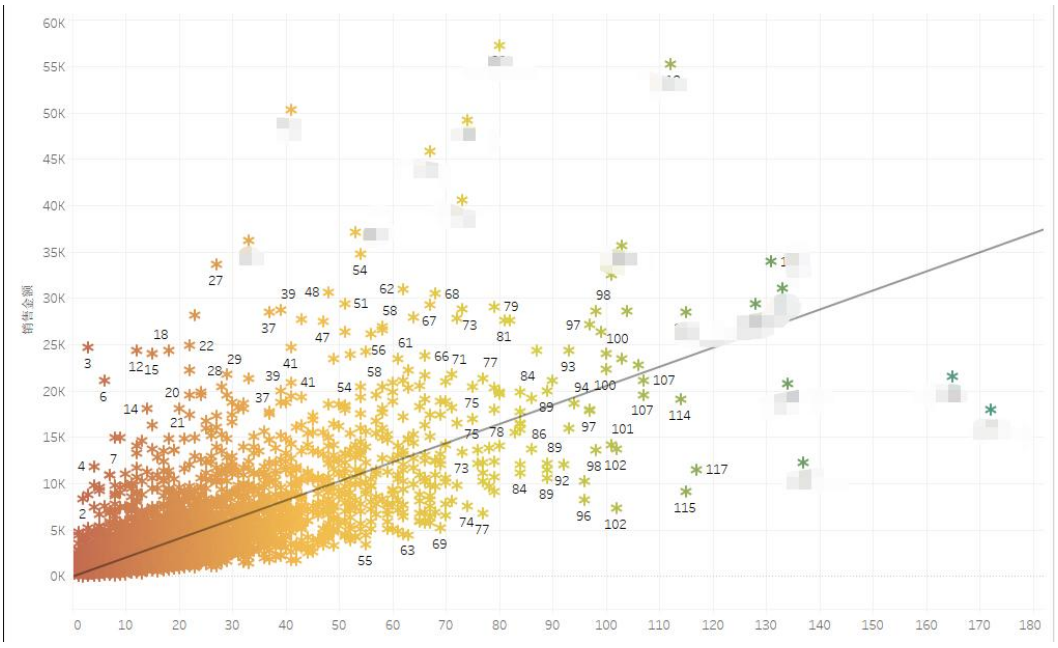
· Set the view size to "Entire View".

· Set the scatter plot color to diverging temperature, with reverse order.

· Set the scatter plot shape to snowflake, consistent with the example.

· Display data labels for the total orders placed per customer.

Reference Image:



icon 10

Module D: Data Development and Application (120 minutes)

Module Description:

As people age, hair loss becomes a significant health concern for many. The fullness of one's hair not only affects appearance but is also closely related to overall health. This dataset compiles various factors that could contribute to hair loss, including genetic factors, hormonal changes, medical conditions, medications, nutritional deficiencies, and psychological stress. Through data exploration and analysis, we can uncover potential correlations between these factors and hair loss, thereby providing valuable insights for personal health management, medical interventions, and the development of related industries.

Task One: Data Overview Presentation

Task Requirements: Design an interactive data overview application using the Streamlit framework. This application should allow users to view different parts of the dataset via a dropdown menu and offer comprehensive features to help users understand the data characteristics deeply.

Requirements and Functional Descriptions:

·When users select "Data Overview," the application should provide a slider that allows users to choose the number of rows to display, showing the first few rows of the dataset.

·Include a checkbox that lets users opt to display the top N rows of data (checked by default).

This interactive application will enable users to easily access an overview of the data, providing reliable support for subsequent data analysis and decision-making.

Task Two: Feature Selection and Data Column Display

Task Requirements:

Design an interactive application feature using the Streamlit framework to implement feature selection and data column display functionality. This application should allow users to select feature columns of the data via a dropdown menu and use checkboxes to control whether to display the data of the selected feature columns.

Requirements and Functional Descriptions:

1.Feature Selection Function:

·Provide a dropdown menu for users to select feature columns of the data.

·The "Id" column should be selected by default.

2.Data Display Function:

·Provide a checkbox to allow users to choose whether to display the data of the selected feature columns.

·By default, the data of the selected feature columns should be displayed.

This application feature will enable users to flexibly view data from different feature columns and control the display of specific feature columns as needed, thereby facilitating data analysis and decision-making.

Task Three: Data Processing and Sorting

Task Requirements:

Design an interactive application feature using the Streamlit framework to implement data processing and sorting functionalities. This application should allow users to select the data column to process via a dropdown menu. Additionally, provide data

sorting options in the sidebar, enabling users to choose the column for sorting and decide whether to sort the data in ascending order.

Requirements and Functional Descriptions:

1.Data Sorting Function:

·Provide data sorting options in the sidebar:

- A dropdown menu for users to select the column to sort.
- The "Id" column should be selected by default.
- A checkbox to allow users to choose ascending order sorting, which should be checked by default.
- A button for users to click to apply the sorting.

2.Sorted Data Display:

·After the user clicks the sorting button, display the data sorted according to the selected column.

This application feature will enable users to easily process and understand specific columns in the dataset, sort the data as needed, and facilitate data analysis and processing.

Task Four: Interactive Analysis of Family History of Hair Loss

Task Requirements:

Design an interactive application feature to dynamically display the proportion of patients with and without a family history of hair loss, visualized through a pie chart.

Requirements and Functional Descriptions:

1.Numerical Display:

·Show the number and percentage of patients with and without a family history of hair loss.

·Display the data for each category in percentage format, rounded to two decimal places.

2.Visualization Options:

·Provide an interactive radio button for users to choose whether to view these proportions through a pie chart.

·If the user selects "Yes," generate a pie chart to dynamically display the proportions of patients with and without a family history of hair loss.

·The pie chart should include detailed data labels showing the specific percentage for each section.

This application feature will enable users to clearly and intuitively understand the distribution of family history data related to hair loss in both numerical and graphical formats, thereby enhancing data analysis and decision-making support.

Task Five: Personal Health Self-Assessment Tool

Task Requirements:

Design a Streamlit application feature that allows users to input information about their personal health status, such as hormonal changes, medical history, and stress levels. Based on this information, the application should provide basic health advice and relevant information resources.

Requirements and Functional Descriptions:

1.Input Section:

·Users input relevant health information through a simple form, including:

- Hormonal changes: Select "Yes" or "No"
- Medical history: Check relevant medical issues from a list

- Stress level: Select "Low," "Medium," or "High"

2.Information Display Section:

·Based on the user's input, display some general health advice as outlined below.

·Provide links or resources related to health information, such as healthy living guides and disease prevention strategies.

3.Health Advice:

Provide the following health advice based on the user's input:

·Hormonal Changes:

- If you are experiencing hormonal changes, it is recommended to undergo regular health check-ups to ensure your hormone levels are within the normal range.

·Medical History:

- For known medical issues, it is advised to follow your doctor's treatment recommendations and have regular medical check-ups to monitor your condition.

·Stress Level:

- If you are feeling high levels of stress, consider trying relaxation techniques such as deep breathing, meditation, or physical exercise to reduce stress and promote mental and physical well-being.

·Lifestyle Habits:

- If you smoke, consider quitting to reduce the risk of cardiovascular diseases and other health issues.
- Maintain a healthy diet and exercise regularly to sustain a healthy lifestyle.

Expected Outcome:

After users complete the health status information, the application feature should provide a page with health advice based on their input, without involving deep analysis.

By offering relevant links and resources, users can gain more information on how to improve or maintain their health.

Task Six: Hair Loss Risk Distribution

Task Requirements:

Design an interactive application feature where users can select their family history of hair loss (Yes/No) via radio buttons and choose their age using a slider. The application should then display the corresponding hair loss risk distribution in a visual chart.

Requirements and Functional Descriptions:

1.User Input Section:

·Family History of Hair Loss: Users select whether they have a family history of hair loss (Yes/No) via radio buttons.

·Age Selection: Users select their age using a slider, with a range set from 18 to 80 years.

2.Visualization Display Section:

·Use a scatter plot to display the distribution of hair loss risk (X-axis: Age; Y-axis: Number). The chart should dynamically update based on the user's family history selection and age range.

After selecting their family history of hair loss and inputting their age, users will see a scatter plot showing the distribution of hair loss risk based on their chosen conditions. This feature enhances the interactivity of data visualization and makes the information easier to understand by allowing users to immediately see the hair loss distribution under different conditions.

Task Seven: Age Distribution Analysis of Patients with Different Medical Histories

Task Requirements:

Design an interactive application feature where users can select one or more specific medical conditions. The application will then analyze and display the age distribution of patients corresponding to the selected medical histories.

Requirements and Functional Descriptions:

1.User Input Section:

·Medical Condition Selection: Users select one or more medical conditions (e.g., alopecia areata, thyroid problems, scalp infections) using checkboxes.

2.Data Analysis and Visualization Section:

·Based on the user's selected medical conditions, the application will analyze the age distribution of patients with those medical histories (X-axis: Age; Y-axis: Number).

·Use a line chart to display the distribution of patients across different age groups.

After selecting specific medical conditions, the application will analyze and display the age distribution of patients with those medical histories. The chart provides a clear visual representation, allowing users to understand the relationship between medical history and age. This helps healthcare professionals and researchers make more accurate judgments about the age distribution of specific diseases.

Task eight: Data Analysis of the Relationship Between Stress Levels and Hair Loss

Task Requirements:

Design an interactive application feature where users input their stress levels, and the application analyzes and displays the data distribution of the relationship between stress levels and hair loss.

Functional Requirements:

1.User Input Section:

·Stress Level Selection: Users select their current stress level (Low, Medium, High) via radio buttons.

2.Data Analysis and Visualization Section:

·Based on the user's selected stress level, the application analyzes the data on the relationship between stress and hair loss.

·Use a bar chart to display the proportion of hair loss at different stress levels.

After selecting a specific stress level, the application will analyze and display the relationship between stress and hair loss based on the user's input. Through the chart, users can intuitively understand the potential impact of different stress levels on hair loss risk. This provides valuable insights, helping users take appropriate stress management measures to mitigate hair loss issues.

4 Project Module Evaluation Criteria

The evaluation criteria reference table 2 for the project module.

Table 2 Evaluation Criteria

Module	Task	Allocation
A	Data Acquisition and Processing(EXCEL)	25
B	Data Analysis and Operations(PYTHON)	25
C	Data Presentation and Sharing(TABLEAU)	25
D	Data Development and Application(STREAMLIT)	25
Total		100

Note: The final interpretation rights of the sample questions belong to the organizing committee.