**CIS 467 final group project (due by Wednesday, December 15 at 11:59 PM).**

**This is a group project (total 400 points). The groups have been created on Blackboard. Please make only one submission per group and put all your team members’ full names into this Word document. Please also submit a Tableau Workbook file .twb into the Final Team Project folder on Blackboard.** **If you cannot attach the Tableau Workbook .twb file to the Final Team Project folder on Blackboard, please email (one email per group) the Tableau Workbook .twb file to me at** [**mlysyako@simon.rochester.edu**](mailto:mlysyako@simon.rochester.edu) **indicating your class section and your team name from Blackboard and all members in the email.**

YOU WILL HAVE 5 ATTEMPTS. Please submit once if possible, and use other attempts if the submitted files in the first attempt were wrong or corrupted or if you find major mistakes in the first attempt. The last attempt will be graded (as long as it is before the deadline).

The script file final\_project\_database\_cis467.sql creates a database which contains the 11 tables (the database schema is below), with transactional data related to some company operations.

**Please check early that you can create the final project database on your machine.**

**Please review Panopto video for week 7 for a description of this dataset.**

**CIS 467 Final Group Project**

**Group P**

**Maheen Ansari**

**Yufeng He**

**Chenxi Hu**

**Lili Wang**

**Qi Zhang**

Timeline

Description automatically generated

Please put all of your work into **this** **single Word doc and also submit a Tableau Workbook file .twb**. Please see instructions for Tableau below in the question 3.

1. **(160 points)** Design and create a data warehouse for the provided database. The decisions about which fields to include and how to aggregate the data are left to you. You do not need to include every single data point from the 11 tables given. Use your judgement as to what will be interesting/useful for the organization. But please make sure that you pull (combine) data from **at least six tables** and compute relevant aggregate statistics. Please compute relevant aggregate statistics for each table that you join. In your queries later in part 2, you may join your Data Warehouse with other tables to answer useful questions. Please see many examples from class lectures and you may adapt those codes for your purpose (for this dataset).

**Submit a screenshot of the first 25 rows of your data warehouse (paste into this Word document) and the SQL code that you used to create it. Please copy and paste your SQL code into this Word document. If you PC does not show 25 rows of data, please submit what you have (i.e., rows you can see on a screenshot) with a comment that you cannot show 25 rows of data. Please add a description of what your Data Warehouse will be tracking for a company.**

The data warehouse will track the orders for a company including order information, supplier and customer information, product information and employee information. The tables joined for this data warehouse include orders, order\_details, customers, employees, suppliers, products, and categories.

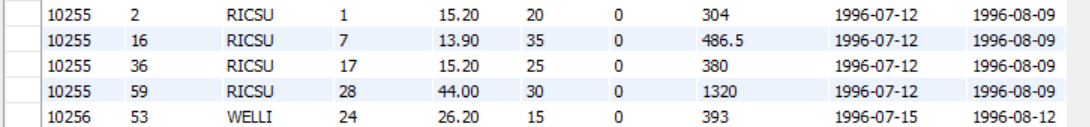
Screenshot of the first 25 rows

**图片包含 表格

描述已自动生成**

**表格

低可信度描述已自动生成**

****

**\*There 35 columns in the data warehouse, the screen cannot show them all.**

Codes

文本

描述已自动生成

文本

描述已自动生成

2. **(140 points)** Create **eight** SQL queries **on your data warehouse** (not on the original dataset) that answer interesting questions. At least **6** queries should be more complex queries. For example, more complex queries could include Joins, a Group By, UNION elements or a subquery or use some aggregate functions and summary calculations (see examples in the class lectures’ slides).

**Submit a copy of each query SQL code (paste into this Word document), and the screenshot of each query results (or the first 25 rows if it is longer or how many rows you can get on your PC) and full description of the question your SQL code was addressing and what you found in the results. The question that each query answers should be useful for a company to make decisions and act upon.**

1. This query provides us with information regarding Order ID, ShipCountry, number of products in that order and order Date. This gives us detailed information of the orders in 1998(the latest year).

Code

文本

低可信度描述已自动生成

The first 25 rows of the result

**表格

描述已自动生成** **表格

描述已自动生成** 图形用户界面, 应用程序

描述已自动生成

1. This query shows columns employeeName, CategoryName and quantityTotal. We are interested in which employee managed the maximum orders on which category. The result shows that e.g.: employee Nancy Dacolio managed the maximum order quantity of 1607 in Beverage category; employee Maegaret Peacock managed the maximum order quantity of 1252 in Condiments category.

Code

文本

描述已自动生成

The first 25 rows of the result

表格

描述已自动生成 表格

低可信度描述已自动生成 图形用户界面, 文本, 应用程序

描述已自动生成

1. This query gives us insight about the number of unique suppliers for each customer and their order. For example: CustomerID ALFKI with OrderID of 10643 has 3 unique suppliers.

Code

图片包含 日程表

描述已自动生成

The first 25 rows of the result

表格

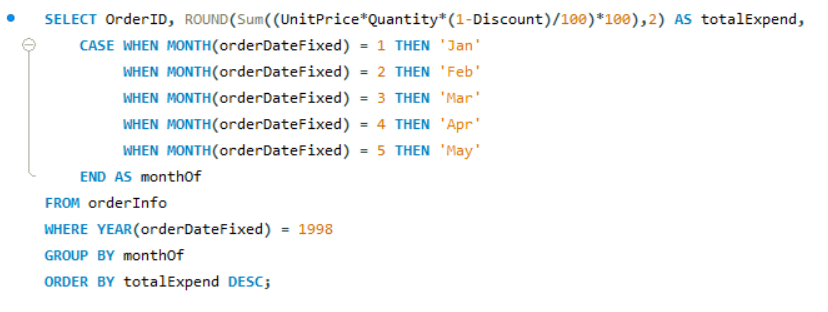
描述已自动生成 图片包含 图形用户界面

描述已自动生成 图形用户界面, 应用程序

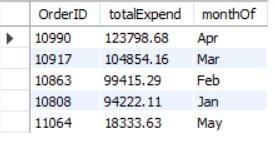
描述已自动生成

1. This query gives us information about the total expend from most to least expend on orders corresponding to months in 1998. For example: In month of April order 11003 had the most expend of 123798.68

Code



Full result



1. The query tells us the date differences between shipdate and required date, which the information contains how many days allowed for the products to arrive its destination. The products belong to certain customer, which the fault\_tolerance\_date is negative or within 10 days need the company's special attention

Code

图形用户界面, 文本, 应用程序

描述已自动生成

The first 25 rows of the result

**表格

描述已自动生成** **表格

描述已自动生成** 图形用户界面, 应用程序

描述已自动生成

1. This query shows the categorized expenditure of top 5 countries. The query compares the expenditure of each category from top 5 countries. For example, in Austria market, dairy products sold the most. (Top 5 countries are defined by the total expenditure in these countries (customerCountry).)

Code

图形用户界面, 文本, 应用程序

描述已自动生成

The first 25 rows of the result

**表格

描述已自动生成** **图形用户界面, 应用程序

描述已自动生成** 图形用户界面, 应用程序

描述已自动生成

1. This query gives us information about the total expend and the number of orders for each customer and label all customers into 5 groups based on the total expend.

Suggestion: Our team suggest the management group to focus more on customers in group 5.

Code

文本

描述已自动生成

The first 25 rows of the result

表格

描述已自动生成 表格

描述已自动生成 图形用户界面, 应用程序, Teams

描述已自动生成

1. This query gives us information about how many orders exceed average shipping time in each year (1996,1997, and 1998) and we calculate the proportion of orders whose shipping time exceeds the average time in total orders. For example, in 1996, in total 152 orders, there were 43 orders have exceeded the average shipping time.

Code

图片包含 图形用户界面

描述已自动生成 文本

描述已自动生成

Full result

图形用户界面, 应用程序

描述已自动生成

3. (**100 points**) Create **five** Tableau individual visualizations (graphs) **on your data warehouse** with valuable information to present findings to senior management of the company. Save each visualization as a png file (as we will practice in the lab 5) and paste each individual visualization png file **into this Word** document with the full explanation of what the visualizations show, how they are useful to a company and how company management could make decisions based on what you show. Finally, combine those **five** visualizations into one **Dashboard** (as we will practice in the lab 5), and save this Dashboard as a png file and **paste the Dashboard into this Word** document.

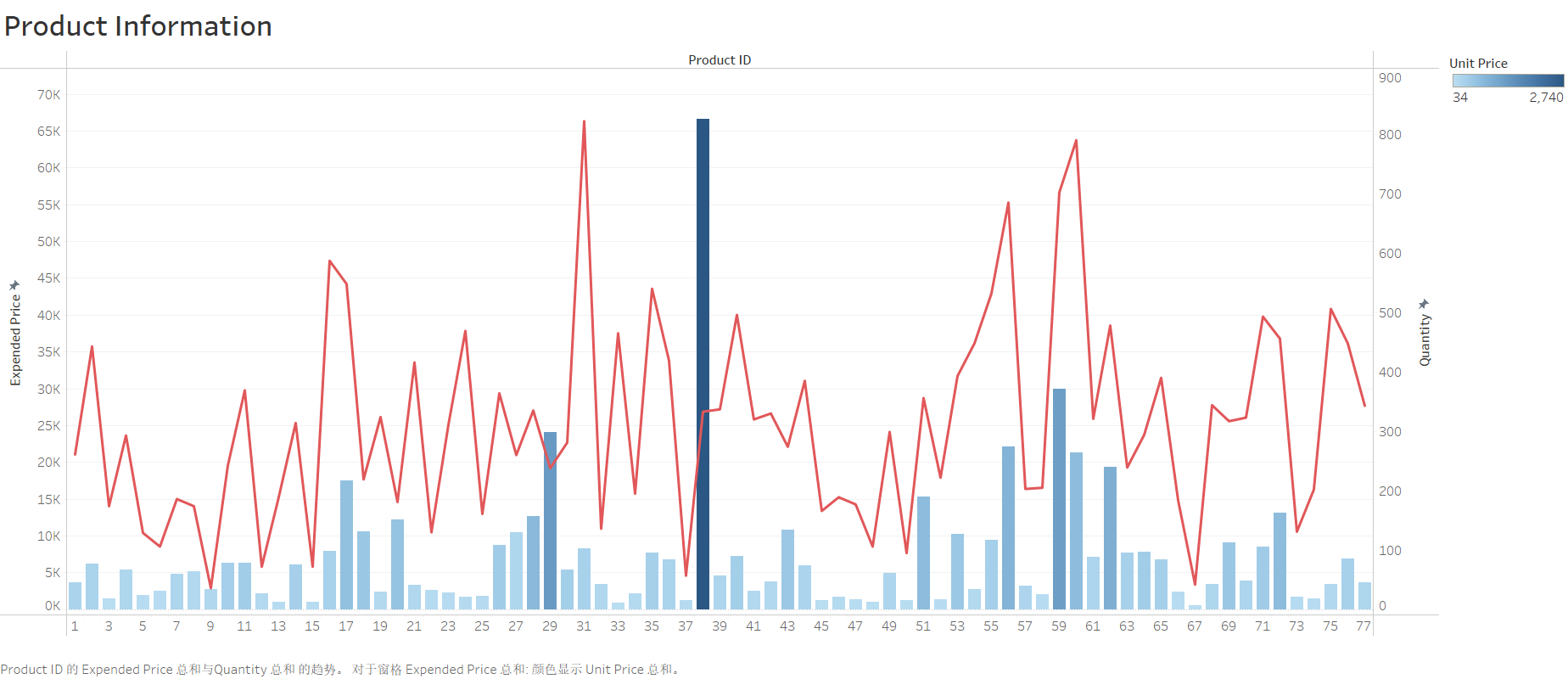
**Please also save the whole Tableau project as a Tableau Workbook file .twb (In Tableau use File - Save as) and submit to the Final Team Project folder on Blackboard together with this Word document. If you cannot attach the Tableau Workbook .twb file to the Final Team Project folder on Blackboard, please email the Tableau Workbook .twb file to me at** [**mlysyako@simon.rochester.edu**](mailto:mlysyako@simon.rochester.edu) **indicating your class section and your team name from Blackboard and all members in the email.**

1. The visualization shows the categorized expenditure of the top 5 countries whose customers have the highest total expenditure. A company can see the rank of each country for each category and design promotional activities of each category for the country that has highest expenditure on that category to improve international trade.

图表, 条形图

描述已自动生成

1. The visualization shows the trend between the unit prices and the quantities sum of products. A company can see product 38 has the highest price and product 31 has the highest quantity. This detailed product information can help tailor the advertising and promotion of products.

****

**General grading criteria: Your completed work will be evaluated using the criteria below. I encourage you to use your creativity and other business skills (communication, presentation, critical thinking) in addition to the data management concepts and the SQL and Tableau skills that we have covered in CIS467.**

|  |  |  |  |
| --- | --- | --- | --- |
| **High score** | **Score between high and good** | **Good/medium score** | **Low score** |
| All required parts of the final project are complete and technically correct. Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries. Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions. | All required parts of the final project are complete and technically correct (with possibly a few minor errors). Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries (with possibly a few minor errors). Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions (with possibly a few minor errors). | Some required parts of the final project are missing and/or there are more significant errors. Some queries appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple but may still provide interesting useful information based on which senior management of the company can make important decisions. | The final project has large portions missing and/or major conceptual errors. Most/all queries (if any) appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple and **do not** provide interesting useful information based on which senior management of the company can make important decisions. |