Problem Statement - New Wheels

Submission type
:
File Upload
Due Date
:
Dec 09, 11:00 PM
Total Marks
:
40
Available from
:
Nov 16, 6:30 PM

Description

Business Context:

A lot of people in the world share a common desire: to own a vehicle. A car or an automobile is seen as an object that gives the freedom of mobility. Many are now preferring pre-owned vehicles because they come at an affordable cost, but at the same time, they are also concerned about whether the after-sales service provided by the resale vendors is as good as the care you may get from the actual manufacturers. New-Wheels, a vehicle resale company, has launched an app with an end-to-end service from listing the vehicle on the platform to shipping it to the customer's location. This app also captures the overall after-sales feedback given by the customer.

Problem Statement:

New-Wheels sales have been dipping steadily in the past year, and due to the critical customer feedback and ratings online, there has been a drop in new customers every quarter, which is concerning to the business. The CEO of the company now wants a quarterly report with all the key metrics sent to him so he can assess the health of the business and make the necessary decisions.

Objective:

As a data scientist, you see that there is an array of questions that are being asked at the leadership level that needs to be answered using data. But this data is not organized and is just being dumped as flat files to be used only when there are concerning problems! You need to first create a pipeline to organize and maintain this data using a SQL database so that it becomes easy to answer questions in the future. Once this is done, use the data to answer the questions posed and create a quarterly business report for the CEO.

Data Description:

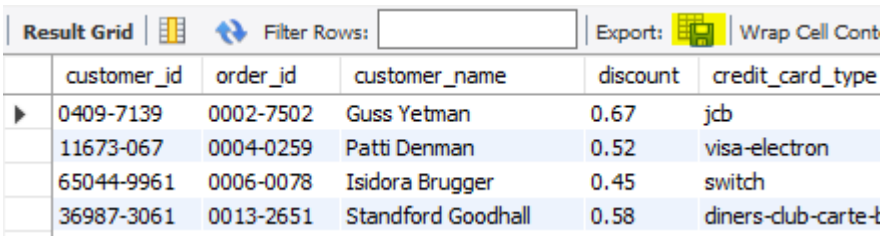
You have data on the vehicles you sell: what is the make, model, and year? What is the price point? Data on your customers, such as where they live and payment methods, data on orders and shipments, such as when the order was shipped and received, what the after-sales feedback was, and so on.

This data is provided for each quarter in the **Data** folder in **new_wheels_proj.zip**. The data dictionary (new_wheels_sales.txt), which has detailed descriptions of each column, can be found in the same **Data** folder.

Solution Approach:

- Follow the same steps that you are familiar with in Week 2: Data Modeling and Architecture. Use the *architecture solution* provided in the **Design** Folder to create the Ingestion, Transaction, and Consumption Layers. Use the *ER Diagram* provided in the same folder to normalize the main table. For creating the views, use the column schema provided in the *View Schema* file. **Note:** Create a new database called **vehdb** to store all the tables, views, procedures, and functions.
- You can find the questions that need to be answered for the business in the project guideline document-submission_myname.*sql*.
- Once these questions have been answered by writing SQL queries, you can use those results to create the Quarterly Business Report. Use the results from the queries* to generate visuals in Python using matplotlib or seaborn libraries. A template report, *new_wheels_sample_QBR_template.pdf,* has been uploaded to this module.

**You can download the results of queries from SQL workbench using the download button as shown below:*



customer_id	order_id	customer_name	discount	credit_card_type
0409-7139	0002-7502	Guss Yetman	0.67	jcb
11673-067	0004-0259	Patti Denman	0.52	visa-electron
65044-9961	0006-0078	Isidora Brugger	0.45	switch
36987-3061	0013-2651	Standford Goodhall	0.58	diners-club-carte-b

Submission Guidelines:

There are three files that are expected in the submission of this project:

Script: This should have the *submission_myname.txt file which contains the below things-*

Note: Use the submission_myname.sql (project guideline document) and write all the codes in the same provided document and save it as a .txt file, called submission_myname.txt, with proper comments for each section.

- DDL scripts for the temp table, the main table, and the normalized tables [*temp_t., vehicles_t, customer_t, order_t, product_t, shipper_t*]
- Stored procedures you created to move data in the database [*vehicles_p*, *customer_p, order_p, product_p, shipper_p*]. ***quarterly_dump is not a procedure, it is the code you use to dump CSV files into the temp table in SQL.*
- Views you have created [*veh_ord_cust_v, veh_prod_cust_v*]
- Functions you have created [*days_to_ship_f, calc_revenue_f*]
- The solution to the questions

Note: All questions must be answered using the views you create in the consumption layer.

Slides: This should have the Quarterly Business Report in the form of a **.pdf**

Python notebook: The notebook should be submitted as an HTML file (**.html**)

Any assignment found copied/plagiarized with other groups will not be graded and will be awarded zero marks.

Please ensure timely submission as any submission post-deadline will not be accepted for evaluation. Submission will not be evaluated if it is submitted post-deadline.

Best Practices for SQL Script:

- The solution script file should be well-documented, with comments (if needed) explaining the functionality of code or the observations and insights.
- It is important to have **no errors** before submission.
- It is recommended that you read the problem statement and go through the criteria and description mentioned in the rubric before starting the project.

Best Practices for Report/Presentation:

- The report should be made keeping in mind that the audience will be the CXOs of the Company or Business Stakeholders.
- The presentation **can** contain additional charts apart from the questions answered in the project.
- The key points in the report should be the following:
    - Business Overview of the problem and solution approach
    - Key findings and insights in each section
    - Business recommendations that can drive decisions
- Focus on explaining the key takeaways in an easy-to-understand manner
- The **Quarterly Report** should be submitted as a **.pdf** file

A business report template has been provided for reference. *It is **not** mandatory to make your presentation in the same style or manner as this sample report. You can get creative!*

Power Ahead!

Scoring guide (Rubric) - SQL and Databases - Project

Criteria	Points
Tables Creation [1] 0.5 pts for DDL code for temp_t [2] 0.5 pts for DDL code for vehicles_t [3] 0.5 pts for DDL code for customer_t [4] 0.5 pts for DDL code for order_t [5] 0.5 pts for DDL code for product_t [6] 0.5 pts for DDL code for shipper_t	3
Stored Procedures Creation [1] 0.5 pts for stored procedure code for vehicles_p [2] 0.5 pts for stored procedure code for customer_p [3] 0.5 pts for stored procedure code for order_p [4] 0.5 pts for stored procedure code for product_p [5] 0.5 pts for stored procedure code for shipper_p	2.5
Data Ingestion [1] 0.5 pts for dump code (quarterly dump)	0.5
Views Creation [1] 0.5 pts for view code for veh_ord_cust_v [2] 0.5 pts for view code for veh_prod_cust_v	1
Functions Creation [1] 0.5 pts for function code for days_to_ship_f [2] 0.5 pts for function code for calc_revenue_f	1
Answering the Business Questions Total 10 questions: [1] Q2, Q3, Q5 and Q7 carry each 2 pts. [2] Q1, Q4, Q6, Q8, Q9 and Q10 carry each 1 pts	14
Plotting Visualizations in Python 0.5 pts for each visuals (total 10 visuals represents each question) generated in Python using Matplotlib or Seaborn for each chart on the Business Report	5
Quarterly Business Report [1] 0.25 pts for each metrics present in overview slide (Total Revenue, Total Orders, Total Customers, Average Rating, Last Quarter Revenue, Last Quarter Orders, Average Days to Ship, % Good Feedback) (0.25 * 8) [2] Overall 2 pts for each observations / findings in each Visual [3] Overall 2 pts for insights and recommendations	6

Criteria		Points
SQL Query Hygiene [1]1.5 pts for well formatted & indented SQL code [2] 1 pts for standard naming conventions are followed [3] 1.5 pts for aliases are given to aggregated columns		4
Report Design Elements Overall 3 marks for following things below- [1] 0.5 pts for standard fonts used in the whole presentation [2] 0.5 pts for charts are cleanly visible with labels [3] 1 pts for findings and observations are written in a crisp manner [4] 1 pts for overall aesthetics and feel of the report (Color choices, Font Sizes, Object Placements and Alignments, etc)		3
	Points	40

Submit Assignment