Data Analyst Assignment Result

# Test Case 1: SQL and Data Manipulation with Python

## SQL

Table Creation Query

|  |
| --- |
| CREATE TABLE customers (  customer\_id VARCHAR(10) PRIMARY KEY  , signup\_date DATE  , last\_purchase\_date DATE  );  CREATE TABLE purchases (  purchase\_id VARCHAR(10) PRIMARY KEY  , customer\_id VARCHAR(10)  , item VARCHAR(100)  , amount INT  , FOREIGN KEY (customer\_id) REFERENCES customers(customer\_id)  ); |

purchases(customer\_id) is referred to customers(customer\_id) since the every purchases is supposed to be made by a customer and cannot have a null value.

1. *Write an SQL query to find the top 10 customers with the highest total purchase amount.*

|  |
| --- |
| SELECT customer\_id, SUM(amount) as total\_purchase\_amount  FROM purchases  GROUP BY 1  ORDER BY 2 DESC  LIMIT 10 |

Result:

|  |
| --- |
| +-------------+-----------------------+  | customer\_id | total\_purchase\_amount |  +-------------+-----------------------+  | 838 | 3376000 |  | 925 | 3293000 |  | 982 | 3282000 |  | 257 | 3274000 |  | 560 | 3256000 |  | 621 | 3214000 |  | 32 | 3214000 |  | 510 | 3155000 |  | 591 | 3154000 |  | 589 | 3119000 |  +-------------+-----------------------+  10 rows in set (0.047 sec) |

1. *Write an SQL query to compute the average purchase amount of customers that have made purchases in the last 30 days*

|  |
| --- |
| SELECT AVG(amount) as avg\_purchase\_amount  FROM purchases  JOIN customers USING (customer\_id)  WHERE last\_purchase\_date >= DATE\_SUB(NOW(), INTERVAL 30 DAY) |

Result

|  |
| --- |
| +---------------------+  | avg\_purchase\_amount |  +---------------------+  | 240333.3333 |  +---------------------+  1 row in set (0.021 sec) |

## Python

Script, result, and comments can be viewed on ‘Test Case 1 - SQL Python.ipynb’ file.

# Test Case 2: Dashboard Creation with Tableau

*Create a Tableau dashboard that:*

1. *Shows monthly sales trends over the 5 years.*
2. *Compares sales by Product Category.*
3. *Provides a regional breakdown of sales.*
4. *Incorporate a feature where a user can select a specific month and view a detailed breakdown of sales for that month.*

*Write a short report detailing:*

1. *Any noticeable trends or patterns in the data.*
2. *Recommendations for the company based on the data.*

*Please make sure to use the correct table format, referring to data warehousing staging, to achieve a responsive and well-performed dashboard.*

The dashboard can be viewed on ‘Test Case 2 - Sales Dashboard.twb’ file.

The file contains 2 dashboards that gives the information of the comparation of each product category in terms of sales trends and breakdown per region. There are date filters, region filters, and product category filters that can be used by user.

Sales figures haven’t been clear needed to be seen to whom. So these figures is made

QRIS or offline payments sales in East Java is the lowest amongst the other regions. It seems that even though QRIS is quite popular in other regions, it is not that popular in East Java. Although, Prepaid Pulsa is the most popular product in East Java. We could optimize the promotion for QRIS in East Java to increase the sales number.

On the other hand, Jabodetabek becomes the least sales region in Java with Send Money as the least sales product in there. Need more data to investigate why is this happening (for example the total number of users since Jabodetabek has the smallest area, maybe we can tell the portion of each region, is Jabodetabek has the most user portion per area?

Every year, our total sales amount is decreasing. We need to create a strategy to increase our sales.

All products have their sales peak in May 2019. Especially for QRIS product, it hasn't been increased since 2019 even though QRIS has been used more and more in society until now. Looks like people tends to use competitors’ QRIS product.

# Test Case 3: Managing Dashboard Lifecycle

*Context:*

*A payment company is planning on using a dashboard to monitor the monthly number of* ***user complaints, treatments given,*** *and* ***average resolution times.***

*Tasks:*

*Outline a plan to design, execute, test, release, and maintain the dashboard. Make sure your answer contains end-to-end workflow for every task needed to create the best dashboard in terms of:*

*1. Data Integrity*

*2. Data Reliability*

*3. Data Security*

*4. User Experience*

*5. Relevance*

*6. Data Completeness*

*7. Performance*

*8. Scalability*

*9. Interactiveness*

*10. Other*

*Suggest three potential ways the data from the dashboard can be used to improve the services of the payment company, ensuring you detail how each suggestion is backed by the data.*

Objectives:

Monitor monthly number of:

* user complaints
* treatments given
* average resolution times

by using dashboard.

Questions to ask:

* Who will be given access to the dashboard? Payment team? Executives? Data team?
* When is the dashboard will be accessed? Every end of the month? Every first week of the month?
* Why do users need to see the data? What kind of metrics that user needed to see?
* How will the user use the data? What is the focus or goal of the user?  
  Is it to increase the average resolution times?   
  Or to increase the number of complaints that is resolved?
* What are the actions that will be done by user if they already got the information from the dashboard?
* What kind of data that we have in existing data warehouse?

Action Plan:

**DESIGN**

1. Do requirement gathering with primary users for initial dashboard requirements. Make sure that all questions to ask is clearly answered by user. If necessary, repeated brainstorming with user can be done for details needed.
2. Explore the existing data. Do we already have all the data? If not, how can we get the missing data?
3. If the data is missing, ask data engineer if they could create a pipeline to the source of data
4. If the data is missing, ask data warehouse developer if they could add the data to the existing warehouse table.
5. Plan the visualization of each information based on the metrics needed and what is the focus.
   1. For example, if user needed to see the trends of user complains per month (or breakdown per week), use line chart to explain the trend itself. Use color code if there's category breakdown needed.
   2. If user needed to see the sum or average number per month, we should use tables, or a big number focused on the top of other visualization as the summary.
   3. If user needed to see funnel of state, use Sankey diagram or bar chart to explain how the user journey of user complaints is.

**DEVELOPMENT**

1. Generate the SQL query that summarizes the monthly data, breakdown by categories, regions, products, etc.
2. Check if the view query is already consistent with existing table/dashboard by checking the general metrics (for example GBV, sales amount, order count). Make sure that the table used in the query is available for production.
3. Create the dashboard based on the design.
4. Check if the dashboard numbers and data is already consistent with the created view (if not, maybe there's inconsistency filters)/existing dashboard by checking the general metrics (for example GBV, sales amount, order count). Visual aspects are also necessary to be considered in this step.
5. Check if the dashboard load time is below the maximum standard.

**TEST**

1. Give primary users access to the draft dashboard.
2. UAT (estimated for a week or two). Make sure that the dashboard experience is good enough for users. Data validation also needed to be checked from user.
3. Fixing the final dashboard if needed from user's feedback.

**PRODUCTION & MAINTENANCE**

1. Official release of the dashboard to primary and secondary users
2. Maintenance if there's any feedback from users
3. Evaluate how is the dashboard usage after release and what’s the impact for the business (especially in customer services) in specific time periods (in a month or a quarter)