**Advanced Data Management – D191**

**VDM1 – VDM1 TASK 1: AUTOMATING DATA INTEGRATION**

**John Funk – 010010682**

Part A - Summarize **one** real-world business report that can be created from the attached Data Sets and Associated Dictionaries.

Business Question

What movie genre do customers spend the most money renting?

Part A1 - Describe the data used for the report.

The data used for this report came from the provided DVD Rental Database. The database contains information about actors, films, inventory, sales, and other topics relating to a DVD rental company. This report uses the category, film\_category, film, inventory, rental and payment tables to compare sales to movie genres.

Part A2 - Identify **two** or more specific tables from the given dataset that will provide the data necessary for the detailed and the summary sections of the report.

This report uses two tables, sales\_genre and sales\_genre\_summary. These tables were created with data from the category, film\_category, film, inventory, rental and payment tables. The sales\_genre table contains the details for every record used in this report. The sales\_genre\_summary table only displays the sum of total sales by movie genre.

Part A3 -  Identify the specific fields that will be included in the detailed and the summary sections of the report.

The table sales\_genre is the detailed table. It uses the following fields.

* genre\_id – This field is the primary key for the sales\_genre table.
* category\_id – This field shows the category ID.
* name – This field shows the name of the category.
* film\_id – This field shows the film ID.
* inventory\_id – This field shows the inventory ID.
* rental\_id – This field shows the rental ID.
* amount – This field shows the amount of each payment in a money format.

The table sales\_genre\_summary is the summary table. It uses the following fields.

* summary\_genre\_id – This field is the primary key for the sales\_genre\_summary table.
* name – This field shows the name of the category.
* total\_sales – This field shows the sum of all sales for each category.

Part A4 - Identify **one** field in the detailed section that will require a custom transformation and explain why it should be transformed. For example, you might translate a field with a value of ‘N’ to ‘No’ and ‘Y’ to ‘Yes’.

This report performs a transformation on the amount field. It takes the amounts of each individual rental in the sales\_genre table and sums the amounts by movie genre in the sales\_genre\_summary table. This allows the user to quickly see which genres produce the most money without scrolling through the entire sales\_genre table.

Part A5 - Explain the different business uses of the detailed and the summary sections of the report.

In addition to answering the business question, the detailed section of this report could also be used to determine which individual movies provide the most sales. Because the total amounts are not summed until we get to the summary table, individual movies could also be examined in the detailed table. The summary table could also be used to find the worst performing movie genres. For example, if comedy movies provide the least amount of sales, the company could decide to reduce the percent of inventory comedy movies occupy.

Part A6 - Explain how frequently your report should be refreshed to remain relevant to stakeholders.

This report should be refreshed monthly to remain relevant to stakeholders. A month will provide enough time for the company to take action based on the results of this report, so refreshing after a month allows the company to monitor the effects of its actions.

Part B - Write a SQL code that creates the tables to hold your report sections.

![Graphical user interface, text, email

Description automatically generated]()

Part C - Write a SQL query that will extract the raw data needed for the Detailed section of your report from the source database and verify the data’s accuracy.

The first screenshot shows the data extraction for the detailed table.

![Graphical user interface, text, application, email

Description automatically generated]()

The second screenshot shows the data extraction for the summary table. I combined the part D transformation with the data extraction for the summary table.

![Graphical user interface, text, application, email

Description automatically generated]()

Part D - Write code for function(s) that perform the transformation(s) you identified in part A4.

This screenshot shows the SUM transformation included in the data extraction for the summary table.

![Graphical user interface, text, application, email

Description automatically generated]()

Part E - Write a SQL code that creates a trigger on the detailed table of the report that will continually update the summary table as data is added to the detailed table.

![Graphical user interface, text, application, email

Description automatically generated]()

![Graphical user interface, text, application

Description automatically generated with medium confidence]()

Part F - Create a stored procedure that can be used to refresh the data in both your detailed and summary tables. The procedure should clear the contents of the detailed and summary tables and perform the ETL load process from part C and include comments that identify how often the stored procedure should be executed.

![Graphical user interface, text, application, email

Description automatically generated]()

![Graphical user interface, text, application, email

Description automatically generated]()

1. Explain how the stored procedure can be run on a schedule to ensure data freshness.

This procedure can be run at any time using the CALL command, so the data science team can manually run the procedure on a specified day each month. The data science team could also write code to automate the process if desired.

Part G -  Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=cd6e7536-104d-4c6d-8233-ae8b01586d6c

Part H - Record the web sources you used to acquire data or segments of third-party code to support the application if applicable. Be sure the web sources are reliable.

The only web source used to acquire data was the provided database from the lab.

Part I - Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

No sources are quoted, paraphrased or summarized.