BeeGeez Sales Dataset Analysis

Notebooks as Business Reports

Due: 23:59, Friday 26th May 2023

BeeGeez is a retail company that operates globally, selling various products through both online and offline sales channels across multiple regions. Over the years, the company has accumulated sales data that includes information on product types, sales channels, regions, revenue, and profit. To improve its sales performance, BeeGeez aims to analyse its sales data and identify factors that contribute to its revenue and profit. Specifically, the company wants to track the evolution of its revenue and profit over time, determine which products and regions generate the most profit, and compare the revenue and profit of its online and offline sales channels.

This assignment aims to analyse BeeGeez's sales dataset to gain insights into various products, regions and sales channels. The analysis will be conducted by leveraging Python programming in Google Colab notebooks, exploring SQLite databases, data visualisation, and applying industry best practices in programming.

Using these insights, BeeGeez can adjust its product mix, pricing strategies, and marketing efforts to increase revenue and profit. Analysing its sales data, Beegeez aims to gain a better understanding of its business performance and make data-driven decisions to optimize its operations and increase its revenue and profit. For instance, the company could focus on selling its most profitable products in regions with high demand or invest more in online marketing to boost online sales revenue.

Assessment Objectives

Analyse the BeeGeez's Sales dataset using Python in Google Colab notebooks to understand SQLite databases, data exploration, data visualization, and applying industry best practices in programming.

Deliver two notebooks: a business report (which includes code, analysis, and discussion) and a development notebook which includes pseudocode, testing, and any other industry best practice not observable from the business report.

Learning Objectives
- Understand and work with SQLite databases.
- Perform data exploration and analysis using Python and relevant libraries.
- Create effective data visualizations.
- Apply industry best practices in programming, such as adding comments, creating a modular design, reusing code, and using version control with GitHub.
- Select and apply appropriate data analysis techniques.
- Interpret and communicate findings effectively.
- Demonstrate critical thinking and problem-solving skills.

The *Business Report Notebook* must run on a Google Colab instance and require no additional steps other than running code cells within the notebook.

Note: You can have code-cells in the notebook set up the Colab instance, for example, copy data, python scripts, or other notebooks. But other than running a code cell your notebook should require no further interaction from the user/reader of the notebook.

Tasks:

- Set up the environment:
  - Create a new Google Colab notebook.
  - Connect the notebook to your GitHub account.
  - Import the necessary libraries (SQLite3, Pandas, Matplotlib, and ipywidgets).

- Access the database:
  - Connect to the BeeGeez Sales Data SQLite database using the SQLite3 library.
  - Examine the schema of the database and understand the structure of the tables.

- Data extraction and manipulation:
  - Write SQL queries to extract relevant information from the tables
  - Use Pandas to load the query results into data frames and perform data manipulation tasks such as filtering, grouping, and aggregation.
  - Clean and pre-process the data, addressing any missing or inconsistent values.

- Interpretation and conclusion:
  - Summarise the main insights you have gained from the data analysis.
  - Discuss any limitations of your analysis and suggest possible improvements.
  - Reflect on the usability and effectiveness of Python notebooks

Analysing and Visualising BeeGeez's Sales Dataset to Gain Valuable Insights

Note that you may need to pre-process and clean the data to ensure accurate results.
For this assignment, you are to include any three with <u>at least one from Group B</u> in your submission.

<u>Group A</u>

Total revenue and total profit over time
Analyse the total revenue and total profit over time by computing the yearly revenues and profits.
Visualisation: Create a time series line chart with the x-axis representing years and the y-axis representing the revenue and profit amount.

Total revenue and profit generated by Product
Analyse the total revenue and profit generated by each item type and then visualise the results sorted in descending order.

Total revenue and profit generated by Region
Analyse the total revenue and profit generated by region and then visualise the results in sorted order.

<u>Group B:</u>

Total revenue and profit generated by Sales Channel
Analyse the total profit and revenue data for Online and Offline sales channels
separately by plotting line graphs for each channel.

Total revenue and profit by sales channel and region
Analyse the total revenue generated by each sales channel and visualise the result as a
pie chart. Next, analyse and visualise total profit by region for each sales channel.

Identifying Top Revenue-Generating Products by Item Type and Sales Channel
Analyse the sales data based on item type and sales channel to identify the products that generate the highest revenue both online and offline. Visualise the results using horizontal bars in either ascending or descending order.

GitHub

Version control is an industry best practice technique for monitoring changes to a file or group of files over time and reverting to a previous version. For this assignment, you are required to create a **new PRIVATE GitHub repository** to store the notebook and any support files. The assignment GitHub repository will contain:

- o   README
- o   Non-Conformance Report
- o   Notebooks required for the assignment
- o   Python scripts required for the assignment
- o   Any other relevant documents

Evaluation

As an IS Professional, you are expected to meet the specification to the best of your ability. This specification is to be treated as the output of a meeting between yourself and a client. Your instructor will take on the role of the client. If you want to implement any functionality or behaviour not described in this specification, please seek approval from the client (*your instructor*) **before** you begin writing your program.

Your submission will be assessed to see if it correctly applies the behaviours mentioned in this document. This problem specification completely describes all behaviours to be tested.

You may only use programming constructs taught in the unit or demonstrated in the textbook. If you plan to use any advanced Python features not introduced in this unit, please seek approval from your instructor **before** you begin writing the program.

The code must follow the programming style naming conventions used in the PEP8, which include:
- Meaningful names for projects, variables, methods, and controls.
- Correct capitalisation of variables and methods
- Appropriate use of comments
- Reference any relevant forums, websites, or videos that you used.
- Use of space and indentation to program is easy to read.

Submission Guidelines

Save your Google Colab notebook(s) as a .ipynb file and push it to your GitHub repository. Write a brief README.md file describing the assignment and the purpose of the repository. Your GitHub repo should be private and contain all documents relevant to this assignment.
Submit the link and zip file to your GitHub repository containing the notebook and README.md file.

This assignment is to be completed individually. **The assignment is due 23:59 Friday 26th May 2023**. The entire assignment GitHub project folder must be submitted as a single compressed archive file to the unit's BlackBoard site submission link.

Non-Conformance Report (NCR)

A non-conformance report (NCR) is a document that addresses issues where there has been a deviation from the project specification or where work fails to meet agreed quality standards. If you cannot implement some functionality or have difficulty meeting any of the requirements, you will need to provide a NCR. An example might be unable to produce the plots, or deviation from the style guide. For each non-conformance issue, you need to document:
- The problem
- Severity and impact
- How it occurred

- How to prevent it from happening again
- Plan or time estimate to fix

Grading Criteria

Your assignment will be graded based on the following criteria:
- Clarity and organization of your code (comments, modular design, code reuse).
- Proper use of version control with GitHub.
- Quality and completeness of the business report (literate programming, clear explanations, and visualizations).
- Effectiveness of the code testing notebook in identifying and resolving issues.
- Overall data analysis quality, including insights and findings based on the Enron Mail dataset.
- Critical thinking and problem solving skills

Academic Integrity

Curtin's Academic Integrity policy must be followed in all submissions. For more details, go to the Academic Integrity tab in Blackboard or the Academic Integrity website. Both submissions must adhere to the Copyright Act of 1968 as well as the 'Digital Agenda' revisions to the Copyright Act.