Problem Statement - Crimes\_LA

**Submission type**

:

File Upload, Text Entry

**Total Marks**

:

40

**Description**

**Problem statement:**

People all throughout Los Angeles are concerned about recent reports of crimes in numerous locations. The mayor of Los Angeles has established a new Criminal Investigation Division to study how and why crime is on the rise, as well as the elements that contribute to it, so that officials may take the necessary steps to keep the city's residents safe.

**Objective**:

As a member of the Analytics Division, you are aware that there are numerous questions that must be answered utilizing CID data. However, this data is not organized and is simply exported as flat files that can only be viewed if there are problems! You must create a pipeline to organize and manage this data using a SQL database to make it much easier to answer inquiries in the future. Then, using the data, react to the questions and write a detailed report for the authorities to utilize in taking action against crimes in Los Angeles.

**Data Description:**

You have information on several types of crimes that occurred in different areas throughout Los Angeles: what was the crime, who was the victim, and where did the incident occur? What is the age of the victim and so on?

This data is provided for each week in the **Data**folder in **crime\_la\_proj.zip**. The data dictionary (crime\_la.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 **crime** 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 Weekly Business Report. Use the results from the queries**\*** to generate visuals in Python using matplotlib or seaborn libraries. A template report, *crime\_la\_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:*

*Table

Description automatically generated*

**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, crime\_t, officer\_t, victim\_t, report\_t, location\_t*]
* Stored procedures you created to move data in the database [*crime\_p,*report\_p*,officer\_p,location\_p,victim\_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 [*rep\_loc\_off\_v and rep\_vict\_v]*
* Functions you have created [*time\_f ,age\_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 Scripts:**

* 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 top government officials.
* 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 **Weekly 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 (1)**

| **Criteria** | **Points** |
| --- | --- |
| **Tables Creation**  [1] 0.5 pts for DDL code for temp\_t [2] 0.5 pts for DDL code for crime\_t [3] 0.5 pts for DDL code for officer\_t [4] 0.5 pts for DDL code for victim\_t [5] 0.5 pts for DDL code for report\_t [6] 0.5 pts for DDL code for location\_t | 3 |
| **Stored Procedures Creation**  [1] 0.5 pts for stored procedure code for crime\_p [2] 0.5 pts for stored procedure code for report\_p [3] 0.5 pts for stored procedure code for officer\_p [4] 0.5 pts for stored procedure code for location\_p [5] 0.5 pts for stored procedure code for victim\_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 rep\_loc\_off\_v [2] 0.5 pts for view code for rep\_vict\_v | 1 |
| **Functions Creation**  [1] 0.5 pts for function code for time\_f [2] 0.5 pts for function code for age\_f | 1 |
| **Answering the Business Questions**  Total 10 questions: [1] Q1, Q3, Q4 and Q8 carry each 2 pts. [2] Q2, Q5, Q6, Q7, 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 |
| **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**

ONLINE UPLOADONLINE TEXT ENTRY

Drag your file(s) here or browse for a file to upload

Add comments

