</talentlabs>

Assignment Chapter 4 - Data Wrangling with SQL

Instructions

1. This assignment is split into 2 parts. For Part 1, no dataset is required. For part 2 you will need to use the boston_crime.csv dataset that was used during the SQL demonstration lessons.
2. Please answer the questions in the boxes provided.
3. Please submit the assignment through the TalentLabs Learning System.

Part 1: SQL Queries

Question 1.1:

Complete the query below to load data without duplicates.

```
SELECT
  DISTINCT *
FROM
  dataset.tableName
```

Question 1.2:

Write a query to select all columns from “cars.database”, and all rows which have missing values in the “mileage” column.

```
SELECT
  *
FROM
  cars.database
WHERE
  mileage IS NULL;
```

Question 1.3:

Following on from question 1.2, write a query to replace the missing values in the mileage column with 0 for rows where the column “condition” has values equal to “new”.

```
UPDATE
  cars.database
SET
  mileage = 0
WHERE
  mileage IS NULL AND condition = 'new';
```

Question 1.4:

Write a query to select 3 columns (“Date”, “Purchase_Price”, “Purchase_Desc”) from the following table: shop.history. Filter the query to only include data for dates (in “Date” column) between Jan 1st 2019 and April 1st 2022. Finally, order the resulting table by the “Purchase_Price” column with the highest value first.

```
SELECT
  Date, Purchase_Price, Purchase_Desc
FROM
  shop.history
WHERE
  Date BETWEEN '2019-01-01' AND '2022-04-01'
ORDER BY
  Purchase_Price DESC;
```

Part 2 – Data Wrangling with SQL

For part 2 of this assignment you will need to use the boston_crime.csv dataset. Make sure your data set id is boston, and the table name is crime (FROM boston.crime).

Question 2.1:

How many entries (rows) does this dataset contain?

The dataset contains 319,073 entries.

Question 2.2:

How many unique offense codes are present within the data? Use the Group By command to find your answer. In the box below, please provide your answer to the question and the query used.

Unique offense codes:

- There are 222 unique offense codes.

Query:

```
SELECT
  COUNT(DISTINCT OFFENSE_CODE)
FROM
  boston.crime;
```

Question 2.3:

Find out how many OFFENSE_DESCRIPTION entries contain the word "ASSAULT" as the first word?
– e.g. ASSAULT - AGGRAVATED - BATTERY

In the box below, please provide your answer to the question and the query used.

OFFENSE_DESCRIPTION entries containing "Assault" as the first word:

- There are 23,567 entries where the OFFENSE_DESCRIPTION contains "ASSAULT" as the first word.

Query:

```
SELECT
  COUNT(*)
FROM
  boston.crime
WHERE
  OFFENSE_DESCRIPTION LIKE 'ASSAULT%';
```

Question 2.4:

Make a new column called TIME which contains the time of the offense from the OCCURRED_ON_DATE column. (Hint: you will need to use the CAST and SUBSTR functions together)

In the box below, please provide the query used as well as a screenshot of the query results containing the new TIME column. The column should look like the one in the Sample Screenshot below.

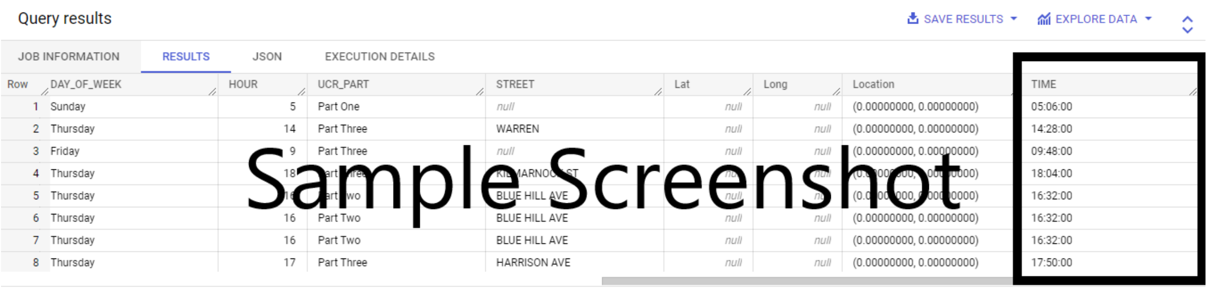
Query:

```
SELECT *,
    CAST(SUBSTR(CAST(OCCURRED_ON_DATE AS STRING), 12, 8) AS TIME) AS TIME
FROM boston.crime;
```

Reasoning:

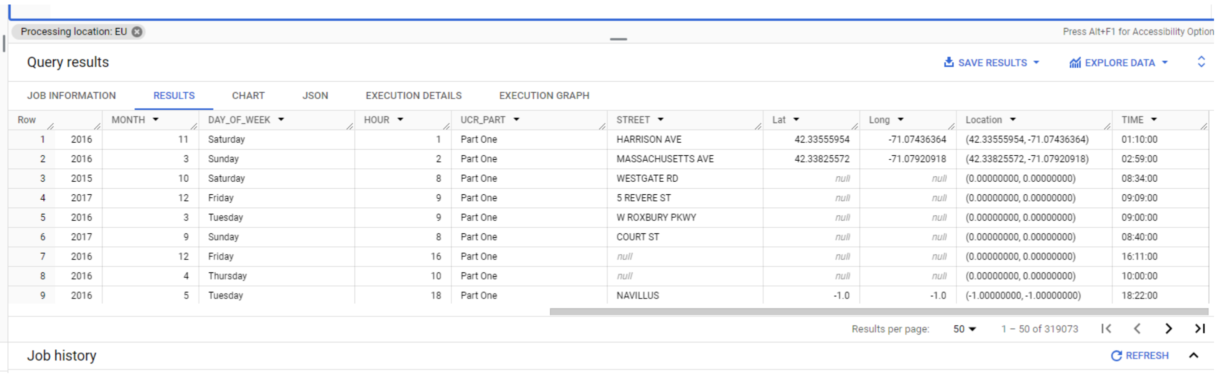
- First change the OCCURRED_ON_DATE to String type, then extract it, starting from the 12th character and 8 characters after that. Then convert it back from String to Time type, and finally name it as a new column named TIME.

Screenshot:



Query results

Row	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location	TIME
1	Sunday	5	Part One	null	null	null	(0.00000000, 0.00000000)	05:06:00
2	Thursday	14	Part Three	WARREN	null	null	(0.00000000, 0.00000000)	14:28:00
3	Friday	9	Part Three	null	null	null	(0.00000000, 0.00000000)	09:48:00
4	Thursday	18	Part Three	KILMARNOCK ST	null	null	(0.00000000, 0.00000000)	18:04:00
5	Thursday	16	Part Two	BLUE HILL AVE	null	null	(0.00000000, 0.00000000)	16:32:00
6	Thursday	16	Part Two	BLUE HILL AVE	null	null	(0.00000000, 0.00000000)	16:32:00
7	Thursday	16	Part Two	BLUE HILL AVE	null	null	(0.00000000, 0.00000000)	16:32:00
8	Thursday	17	Part Three	HARRISON AVE	null	null	(0.00000000, 0.00000000)	17:50:00



Query results

Row		MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location	TIME
1	2016	11	Saturday	1	Part One	HARRISON AVE	42.33555954	-71.07436364	(42.33555954, -71.07436364)	01:10:00
2	2016	3	Sunday	2	Part One	MASSACHUSETTS AVE	42.33825572	-71.07920918	(42.33825572, -71.07920918)	02:59:00
3	2015	10	Saturday	8	Part One	WESTGATE RD	null	null	(0.00000000, 0.00000000)	08:34:00
4	2017	12	Friday	9	Part One	5 REVERE ST	null	null	(0.00000000, 0.00000000)	09:09:00
5	2016	3	Tuesday	9	Part One	W ROXBURY PKWY	null	null	(0.00000000, 0.00000000)	09:00:00
6	2017	9	Sunday	8	Part One	COURT ST	null	null	(0.00000000, 0.00000000)	08:40:00
7	2016	12	Friday	16	Part One	null	null	null	(0.00000000, 0.00000000)	16:11:00
8	2016	4	Thursday	10	Part One	null	null	null	(0.00000000, 0.00000000)	10:00:00
9	2016	5	Tuesday	18	Part One	NAVILLUS	-1.0	-1.0	(-1.00000000, -1.00000000)	18:22:00

Results per page: 50   1 – 50 of 319073

Job history