-- Databricks notebook source

-- MAGIC

-- MAGIC %md-sandbox

-- MAGIC

-- MAGIC <div style="text-align: center; line-height: 0; padding-top: 9px;">

-- MAGIC <img src="https://databricks.com/wp-content/uploads/2018/03/db-academy-rgb-1200px.png" alt="Databricks Learning" style="width: 600px">

-- MAGIC </div>

-- COMMAND ----------

-- MAGIC %md

-- MAGIC # Queries in Spark SQL

-- MAGIC ## Module 1 Assignment

-- MAGIC

-- MAGIC ## ![Spark Logo Tiny](https://files.training.databricks.com/images/105/logo\_spark\_tiny.png) In this assignment you:

-- MAGIC \* Create a table

-- MAGIC \* Write SQL queries

-- MAGIC

-- MAGIC

-- MAGIC For each \*\*bold\*\* question, input its answer in Coursera.

-- COMMAND ----------

-- MAGIC %run ../Includes/Classroom-Setup

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Working with Incident Data

-- MAGIC

-- MAGIC For this assignment, we'll be using a new dataset: the [SF Fire Incident](https://data.sfgov.org/Public-Safety/Fire-Incidents/wr8u-xric) dataset. It has been mounted for you using the script above. The path to this dataset is as follows:

-- MAGIC

-- MAGIC `/mnt/davis/fire-incidents/fire-incidents-2016.csv`

-- MAGIC

-- MAGIC In this assignment, you will read the dataset and perform a number of different queries.

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ## ![Spark Logo Tiny](https://files.training.databricks.com/images/105/logo\_spark\_tiny.png) Create a Table

-- MAGIC

-- MAGIC Create a new table called `fireIncidents` for this dataset. Be sure to use options to properly parse the data.

-- COMMAND ----------

CREATE TABLE IF NOT EXISTS fireIncidents

USING csv

OPTIONS (

header "true",

path "/mnt/davis/fire-incidents/fire-incidents-2016.csv",

inferSchema "true"

)

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Question 1

-- MAGIC

-- MAGIC Return the first 10 lines of the data. On the Coursera platform, input the result to the following question:

-- MAGIC

-- MAGIC \*\*What is the first value for "Incident Number"?\*\*

-- COMMAND ----------

SELECT \* FROM fireIncidents LIMIT 10;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ## ![Spark Logo Tiny](https://files.training.databricks.com/images/105/logo\_spark\_tiny.png) `WHERE` Clauses

-- MAGIC

-- MAGIC A `WHERE` clause is used to filter data that meets certain criteria, returning all values that evaluate to be true.

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Question 2

-- MAGIC

-- MAGIC Return all incidents that occurred on Conor's birthday in 2016. For those of you who forgot his birthday, it's April 4th. On the Coursera platform, input the result to the following question:

-- MAGIC

-- MAGIC \*\*What is the first value for "Incident Number" on April 4th, 2016?\*\*

-- MAGIC

-- MAGIC \*\*Remember to use backticks (\`\`) instead of single quotes ('') for columns that have spaces in the name. \*\*

-- COMMAND ----------

SELECT `Incident Number`,`Incident Date` from fireIncidents

WHERE `Incident Date` = '04/04/2016' ;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Question 3

-- MAGIC

-- MAGIC Return all incidents that occurred on Conor's \_or\_ Brooke's birthday. For those of you who forgot her birthday too, it's `9/27`.

-- MAGIC

-- MAGIC \*\*Is the first fire call in this table on Brooke or Conor's birthday?\*\*

-- COMMAND ----------

SELECT `Incident Date`, `Call Number`

FROM fireIncidents

WHERE `Incident Date`= '04/04/2016' OR `Incident Date` = '09/27/2016' ;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Question 4

-- MAGIC Return all incidents on either Conor or Brooke's birthday where the `Station Area` is greater than 20.

-- MAGIC

-- MAGIC \*\*What is the "Station Area" for the first fire call in this table?\*\*

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ## ![Spark Logo Tiny](https://files.training.databricks.com/images/105/logo\_spark\_tiny.png) Aggregate Functions

-- MAGIC

-- MAGIC Aggregate functions compute a single result value from a set of input values. Use the aggregate function `COUNT` to count the total records in the dataset.

-- COMMAND ----------

SELECT `Incident date`,`Call Number`,`Station Area`

FROM fireIncidents

WHERE `Station Area` > 20 ;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Question 5

-- MAGIC

-- MAGIC Count the incidents on Conor's birthday.

-- MAGIC

-- MAGIC \*\*How many incidents were on Conor's birthday in 2016?\*\*

-- COMMAND ----------

SELECT COUNT (`Incident Number`)

FROM fireIncidents

WHERE `Incident Date` ='04/04/2016';

-- COMMAND ----------

-- MAGIC %md-sandbox

-- MAGIC ### Question 6

-- MAGIC

-- MAGIC Return the total counts by `Ignition Cause`. Be sure to return the field `Ignition Cause` as well.

-- MAGIC

-- MAGIC <img alt="Hint" title="Hint" style="vertical-align: text-bottom; position: relative; height:1.75em; top:0.3em" src="https://files.training.databricks.com/static/images/icon-light-bulb.svg"/>&nbsp;\*\*Hint:\*\* You'll have to use `GROUP BY` for this

-- MAGIC

-- MAGIC \*\*How many fire calls had an "Ignition Cause" of "4 act of nature"?\*\*

-- COMMAND ----------

SELECT `Ignition Cause`,COUNT( `Call Number`) FROM fireIncidents

WHERE `Ignition Cause` = '4 act of nature'

GROUP BY `Ignition Cause` ;

-- COMMAND ----------

-- MAGIC %md-sandbox

-- MAGIC ## ![Spark Logo Tiny](https://files.training.databricks.com/images/105/logo\_spark\_tiny.png) Sorting

-- MAGIC

-- MAGIC ### Question 7

-- MAGIC

-- MAGIC Return the total counts by `Ignition Cause` sorted in ascending order.

-- MAGIC

-- MAGIC <img alt="Hint" title="Hint" style="vertical-align: text-bottom; position: relative; height:1.75em; top:0.3em" src="https://files.training.databricks.com/static/images/icon-light-bulb.svg"/>&nbsp;\*\*Hint:\*\* You'll have to use `ORDER BY` for this.

-- MAGIC

-- MAGIC \*\*What is the most common "Ignition Cause"? (Put the entire string)\*\*

-- COMMAND ----------

SELECT `Ignition Cause`,COUNT(`Ignition Cause`)

FROM fireIncidents

GROUP BY `Ignition Cause`

ORDER BY `Ignition Cause` asc;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC Return the total counts by `Ignition Cause` sorted in descending order.

-- COMMAND ----------

SELECT `Ignition Cause`,COUNT(`Ignition Cause`)

FROM fireIncidents

GROUP BY `Ignition Cause`

ORDER BY `Ignition Cause` desc;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ## ![Spark Logo Tiny](https://files.training.databricks.com/images/105/logo\_spark\_tiny.png) Joins

-- MAGIC

-- MAGIC Create the table `fireCalls` if it doesn't already exist. The path is as follows: `/mnt/davis/fire-calls/fire-calls-truncated-comma.csv`

-- COMMAND ----------

CREATE TABLE IF NOT EXISTS fireCalls

USING csv

OPTIONS (

header "true",

path "/mnt/davis/fire-calls/fire-calls-truncated-comma.csv",

inferSchema "true")

-- COMMAND ----------

-- MAGIC %md

-- MAGIC Join the two tables on `Battalion` by performing an inner join.

-- COMMAND ----------

select \* from fireCalls

inner join fireIncidents

on fireCalls.`Battalion`=fireIncidents.`Battalion`

-- COMMAND ----------

-- MAGIC %md

-- MAGIC ### Question 8

-- MAGIC

-- MAGIC Count the total incidents from the two tables joined on `Battalion`.

-- MAGIC

-- MAGIC \*\*What is the total incidents from the two joined tables?\*\*

-- COMMAND ----------

SELECT COUNT(fireIncidents.`Incident Number`)

FROM fireCalls

INNER JOIN fireIncidents ON

firecalls.`Battalion`=fireIncidents.`Battalion`;

-- COMMAND ----------

-- MAGIC %md

-- MAGIC Congratulations! You made it to the end of the assignment!

-- COMMAND ----------

-- MAGIC %md-sandbox

-- MAGIC &copy; 2021 Databricks, Inc. All rights reserved.<br/>

-- MAGIC Apache, Apache Spark, Spark and the Spark logo are trademarks of the <a href="http://www.apache.org/">Apache Software Foundation</a>.<br/>

-- MAGIC <br/>

-- MAGIC <a href="https://databricks.com/privacy-policy">Privacy Policy</a> | <a href="https://databricks.com/terms-of-use">Terms of Use</a> | <a href="http://help.databricks.com/">Support</a>