الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY

CODING DOJO

FINAL PROJECT /PIG LATIN
TRIPADVISOR EUROPEAN RESTAURANTS

Question templet

DONE BY: DAQUEST GROUP

Q1

Split the dataset based on cuisines (french or Italian) and store each one in a separate dataset

```
-- load the data
-- Split data for cuisines
SPLIT data INTO frensh_cuisines IF cuisines MATCHES '.*French.*', italian_cuisines IF cuisines MATCHES '.*Italian.*';
-- Save the results in seperate datasets
STORE frensh_cuisines INTO '/user/maria_dev/capston project/q3/frensh_cuisines' USING PigStorage (',');
STORE italian_cuisines INTO '/user/maria_dev/capston project/q3/italian_cuisines' USING PigStorage (',');
```

Q2

List the France restaurants that are vegetarian and vegan and have food ratings greater than 4.5

```
-- load the data
-- Take columns we need
columns = FOREACH data GENERATE restaurant_name,country, vegetarian_friendly, vegan_options, food;
-- Filter the data
filter_1 = FILTER columns BY (country MATCHES 'France');
filter_2 = FILTER filter_1 BY (vegetarian_friendly == 1);
filter_3 = FILTER filter_2 BY (vegan_options == 1);
filter_4 = FILTER filter_3 BY (food > 4.5);
-- list the restaurant names
restaurants_list = FOREACH filter_4 GENERATE restaurant_name;
-- Dumb the results
DUMP restaurants_list;
```

Q3

List the top 10 cities and order them by total reviews and save the results

```
-- load the data
--select the needed columns
col = FOREACH data GENERATE city,total_reviews_count;
-- Order the results by total reviews
order_by_data = ORDER col BY total_reviews_count DESC;
--Limit the charges to only the top 10 charges
result = LIMIT order_by_data 10;
--Display the results
DUMP result;
--store the results
STORE result INTO'/user/maria_dev/result/3que.txt'USING PigStorage (',');
```

Q4

Find the Maximum, Minimum total reviews of restaurants located in France and have an average rating of 4.0, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE country,total_reviews_count,avg_rating;
-- filter the column
filter_1 = FILTER col BY (country == 'France') AND (avg_rating == 4.0);
-- group the result
grp = group filter_1 ALL;
-- find the max and min for review count
result = FOREACH grp GENERATE MAX(filter_1.total_reviews_count) , MIN
(filter_1.total_reviews_count);
--Display the results
DUMP result;
STORE result INTO'/user/maria_dev/result/4_que.txt'USING PigStorage (',');
```

Q5

Count the restaurants that are gluten-free and have average ratings higher than 4 in each country, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE restaurant_name,gluten_free,avg_rating,country;
-- filter the column
filter_1 = FILTER col BY (gluten_free == 1) AND (avg_rating > 4.0);
-- group by country
grp = group filter_1 by country;
-- count the restaurant name
result = FOREACH grp GENERATE COUNT(filter_1.restaurant_name);
--Display the results
DUMP result;
STORE result INTO'/user/maria_dev/result/5_que.txt'USING PigStorage (',');
```

Q6

Find the maximum and minimum food ratings of the restaurants that are located in Italy and vegan friendly, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE vegan_options,food,country;
-- filter the column
filter_1 = FILTER col BY (vegan_options == 1) AND (country == 'Italy');
-- group the result
grp = group filter_1 ALL;
-- find the max and min for review count
result = FOREACH grp GENERATE   MAX(filter_1.food) , MIN (filter_1.food);
--Display the results
DUMP result;
STORE result INTO'/user/maria_dev/result/6_que.txt'USING PigStorage (',');
```

Q7

Find the total reviews in English for each city that are located in Greece, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE total_reviews_count,default_language,country,city;
-- filter the column
filter_1 = FILTER col BY (default_language == 'English') AND (country == 'Greece');
-- group the filter by city
grp = group filter_1 by city;
-- Count the review count
result = FOREACH grp GENERATE group , COUNT(filter_1.total_reviews_count);
--Display the results
DUMP result;
STORE result INTO'/user/maria_dev/result/7que.txt'USING PigStorage (',');
```

Q8

Find the average rating of the atmosphere of all the restaurants that are located in Poland and gluten-free

```
-- load the data
--  select column
col = FOREACH data GENERATE atmosphere,country,gluten_free,restaurant_name;
-- filter the column
filter_1 = FILTER col BY (gluten_free == 1) AND (country == 'Poland');
-- group the result by restaurant name
grp = group filter_1  ALL;
-- find the averge of atmosphere
result = FOREACH grp GENERATE   AVG(filter_1.atmosphere);
--Display the results
DUMP result;
```

Q9

Find the maximum reviews that have all languages as the default language of the restaurants that have an excellent rating higher that one thousand , and save the results

```
-- load the data
-- select the column
colm = FOREACH data GENERATE excellent,total_reviews_count,default_language;
-- filter the data by the cuisines and the average ratings
filtering = FILTER colm BY (excellent >= 1000) AND (default_language == 'All languages');
-- Group the filter  by all
grp = GROUP filtering  all ;
-- find the maximum results
results = FOREACH grp GENERATE  MAX(filtering.total_reviews_count);
-- dump the result
DUMP results;
-- store the result
STORE results INTO'/user/maria_dev/final_project_results/Q_9'USING PigStorage (',')
```

Q10

Lists the restaurant name that is located in Spain and vegan and has a cheap prices, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE
restaurant_name,vegan_options,price_range,vegetarian_friendly,country;
-- filter the column
filter_1 = FILTER col BY (vegan_options == 1) AND (price_range == 'cheap') AND (country
== 'Spain');
-- group the result
grp = group filter_1 by restaurant_name;
-- count the restaurant name
results = FOREACH grp GENERATE group ,  COUNT(filter_1.restaurant_name);
--Display the results
DUMP results;
STORE results INTO'/user/maria_dev/final_project_results/Q-10'USING PigStorage (',')
```

Q11

Find the number of restaurants that are located in England and have mid-price range in each city, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE restaurant_name,price_range,country,city;
-- filter the column
filter_1 = FILTER col BY  (price_range == 'mid') AND (country == 'England');
-- group the result
grp = group filter_1 by city;
-- Count the restaurant name
results = FOREACH grp GENERATE group ,  COUNT(filter_1.restaurant_name);
--Display the results
DUMP results;
STORE results INTO'/user/maria_dev/final_project_results/Q-11'USING PigStorage (',')
```

Q12

Count the restaurants that are vegetarian and located in Paris of each price range, and save the results

```
-- load the data
-- select column
col = FOREACH data GENERATE restaurant_name,price_range,city,vegetarian_friendly;
-- filter the column
filter_1 = FILTER col BY  (vegetarian_friendly == 1) AND (city == 'Paris');
-- group the result by price_range
grp = group filter_1 by price_range;
-- count the restaurant name
results = FOREACH grp GENERATE group ,  COUNT(filter_1.restaurant_name);
--Display the results
DUMP results;
STORE results INTO'/user/maria_dev/final_project_results/Q-12'USING PigStorage (',')
```

Q13

List the restaurants that are located in London Colney and have a mid-price range, and order them by the Avg ratings , and save the results

```
-- load the data
-- select the column
colm = FOREACH data GENERATE city,price_range,avg_rating;
-- filter the data by the cuisines and the average ratings
filtering1 = FILTER colm BY (city=='London Colney') AND (price_range== 'mid');
-- Order the filtering data
ordering = ORDER filtering1 BY avg_rating DESC;
-- dump the result
DUMP ordering;
-- store the result
STORE ordering INTO'/user/maria_dev/final_project/Q_13.txt'USING PigStorage (',')
```

Q14

List the restaurants that are located in Spain and average ratings = 5, and service rating= 4.5, and save the results

```
-- load the data
-- select the column
colm = FOREACH data GENERATE restaurant_name,country,service,avg_rating;
-- filter the data by the cuisines and the average ratings
filtering = FILTER colm BY (country =='Spain') AND (avg_rating== 5) AND (service == 4.5);
-- limit the filtering data
lim_results = LIMIT filtering 20;
-- dump the result
DUMP lim_results;
-- store the result
STORE lim_results INTO'/user/maria_dev/final_project_results/Q_14'USING PigStorage (',')
```

Q15

What is the cuisines for the most popularity generic restaurant in Elne

```
-- Take columns we need
columns = FOREACH data GENERATE restaurant_name, cuisines, popularity_generic;
-- Filter data using conditions
filtering = FILTER columns BY (popularity_generic MATCHES '.*#1 .*' AND
popularity_generic MATCHES '.* Elne *');
-- Take the cuisines for the resturant
cuisine =  FOREACH filtering GENERATE cuisines;
-- DUMP to show results
DUMP cuisine;
```

Q16

Count the Italian cuisine restaurants based on each price range

```
--load the data
-- select column
col = FOREACH data GENERATE price_range,city,cuisines;
-- filter the column
filter_1 = FILTER col BY (cuisines == 'Italian' );
-- group the result by price_range
grp = group filter_1 by price_range;
-- count the restaurant name
results = FOREACH grp GENERATE group ,  COUNT(filter_1);
--Display the results
DUMP results;
```

Q17

Count the claimed and unclaimed restaurants per country and save them on a separate file

```
-- load the data
-- Split data for claimed and unclaimed
SPLIT data INTO claimed_restaurants IF claimed == 'Claimed', unclaimed_restaurants IF
claimed == 'Unclaimed';
-- Save the results in separate datasets
STORE claimed_restaurants INTO '/user/maria_dev/final_project_results/Q_11/claimed'
USING PigStorage (',');
STORE unclaimed_restaurants INTO
'/user/maria_dev/final_project_results/Q_11/unclaimed' USING PigStorage (',');
-- Select the needed columns
columns = FOREACH data GENERATE claimed,restaurant_name, country;
-- Group the filtered data by the countries
country_group = GROUP columns BY country;
-- Foreach group count the claimed and unclaimed
claimed_count = FOREACH country_group {
claimed = FILTER columns BY claimed == 'Claimed';
unclaimed = FILTER columns BY claimed == 'Unclaimed';
GENERATE group, COUNT(claimed) as claimed_cnt, COUNT(unclaimed) as
unclaimed_cnt;
};
-- Show the results
DUMP claimed_count;
```

Q18

Find the number of restaurants that have food, service, value, and atmosphere ratings higher than 4, and have mid-price range in each country.

```
-- load the data
col = FOREACH data GENERATE
price_range,atmosphere,cuisines,restaurant_name,service ,country;
-- filter the column
filter_1 = FILTER col BY (service >4.0)AND(atmosphere >4.0)AND(price_range == 'mid');
-- make the group to  generate count
groupeded = Group filter_1 by country;
-- count the restaurant name
results= FOREACH groupeded GENERATE group , COUNT(filter_1);
--Display the results
DUMP results;
```

Q19

Count the restaurants that open more than 150 hours per week and have average ratings equal to 5 in each country

```
-- select column
col = FOREACH data GENERATE open_hours_per_week,avg_rating,country;
-- filter the column
filter_1 = FILTER col BY (avg_rating == 5) AND (open_hours_per_week >=150);
-- make the group for each country to  generate count
groupeded = Group filter_1 BY country;
-- count the restaurant name
results= FOREACH groupeded GENERATE group, COUNT(filter_1);
--Display the results
DUMP results;
```

Q20

Find the average for food rating for all restaurants where the price range are cheap

```
--select the columns
col = FOREACH data GENERATE food,price_range;
-- filter the data using the price range column
filtering = FILTER col BY (price_range == 'cheap');
-- group the filtered data using the country column
grouping = Group filtering all;
-- count the restaurant name
results= FOREACH grouping GENERATE AVG(filtering.food);
--display the results
DUMP results;
```