**PIG QUESTIONS**

TripAdvisor European restaurants

1. Find the top 10 total reviews count based on city

------------------------------------------------------------------------------------------------------------------------------------------

1. Find the Max, Min of total reviews for cuisine frensh and average rating 4.0

------------------------------------------------------------------------------------------------------------------------------------------

3- split the dataset based on cusions (frensh or italian) and stor each one in a seperate dataset

-- 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 (',');

------------------------------------------------------------------------------------------------------------------------------------------

1. Top 10 restaurant for those cuision frensh and AVG rating = 4.5

------------------------------------------------------------------------------------------------------------------------------------------

1. Top 5 resturant in frensh country and avg rating = 5 , service = 4.5

------------------------------------------------------------------------------------------------------------------------------------------

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

------------------------------------------------------------------------------------------------------------------------------------------

1. How many restaurants that gluten-free and have Avg rating higher than 4 in each country?

-- load data and declare the columns data types

data = LOAD '/user/maria\_dev/capston project/cleaning\_data.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage

(',', 'NO\_MULTILINE', 'UNIX', 'SKIP\_INPUT\_HEADER')

as (restaurant\_name: chararray ,country: chararray,region: chararray,city: chararray,latitude: float,longitude: float,claimed: chararray,

popularity\_detailed: chararray,popularity\_generic: chararray,cuisines: chararray,vegetarian\_friendly: int,vegan\_options: int,gluten\_free: int,

open\_days\_per\_week: int,open\_hours\_per\_week: float,working\_shifts\_per\_week: int,avg\_rating: float,total\_reviews\_count: int,default\_language: chararray,

reviews\_count\_in\_default\_language: int,excellent: float,very\_good: float,average: float,poor: float,terrible: float,food: float,service: float,value: float,

atmosphere: float,price\_range: chararray);

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

-- Take columns we need

columns = FOREACH data GENERATE restaurant\_name,country, vegetarian\_friendly, vegan\_options, food;

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);

restaurants\_list = FOREACH filter\_4 GENERATE restaurant\_name;

DUMP restaurants\_list;

------------------------------------------------------------------------------------------------------------------------------------------

1. Find the maximum and minimum food ratings of the restaurants that are located in Italy, and serve pizza.

------------------------------------------------------------------------------------------------------------------------------------------

1. Find the total reviews in English for each city that are located in Greece

------------------------------------------------------------------------------------------------------------------------------------------

1. Find the Avg atmosphere rating of the restaurants located in central Poland and gluten-free

------------------------------------------------------------------------------------------------------------------------------------------

1. Find the Maximum reviews that have all languages as the default of the restaurants that have an excellent rating higher than one thousand