Track Three - Pig Latin
Capstone Project Two
Educational landscape in the MENA region
Desert Ninjas

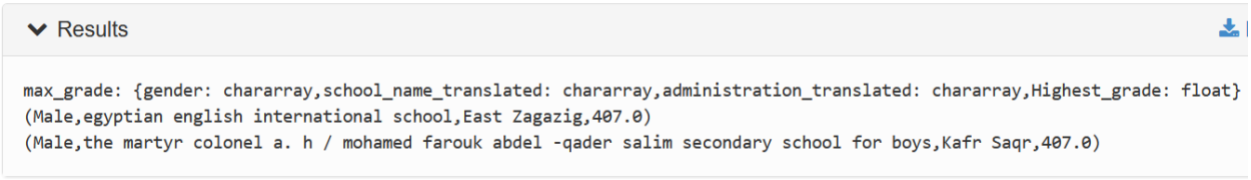
Q1. What is the gender, school name, and administration of the person who got the highest total?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data based on the highest total
filtered = FILTER Egyption_school BY (total>=407);

--3. Grouping the data.
grouped_data = GROUP filtered BY (gender, school_name_translated,
administration_translated);

--4. Get the maximum total.
max_grade = FOREACH grouped_data GENERATE FLATTEN(group) AS (gender,
school_name_translated, administration_translated), MAX(filtered.total) AS
Highest_grade;

--5. Describe the results.
DESCRIBE max_grade;
```

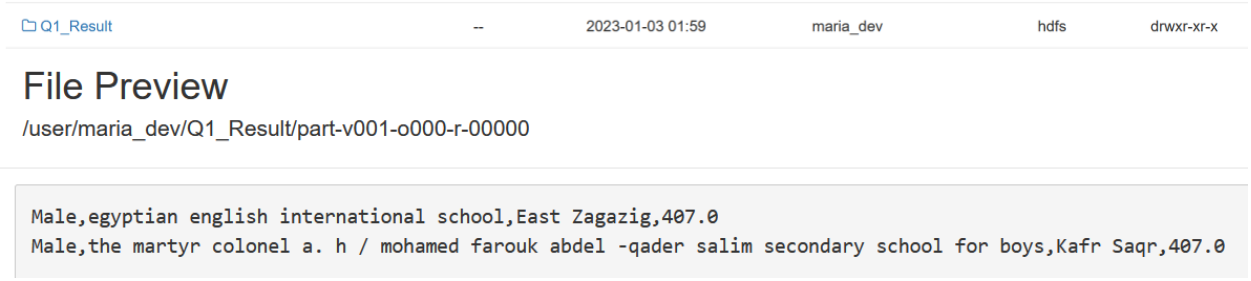
```
--6. Display the results.
DUMP max_grade;

--7. Store the results.
STORE max_grade INTO '/user/maria_dev/Q1_Result' USING PigStorage (',');
```

- **The Result**

max_grade: {gender: chararray,school_name_translated: chararray,administration_translated: chararray,Highest_grade: float}
(Male,egyptian english international school,East Zagazig,407.0)
(Male,the martyr colonel a. h / mohamed farouk abdel -qader salim secondary school for boys,Kafr Saqr,407.0)



- **Storing result**



- **Insight:**

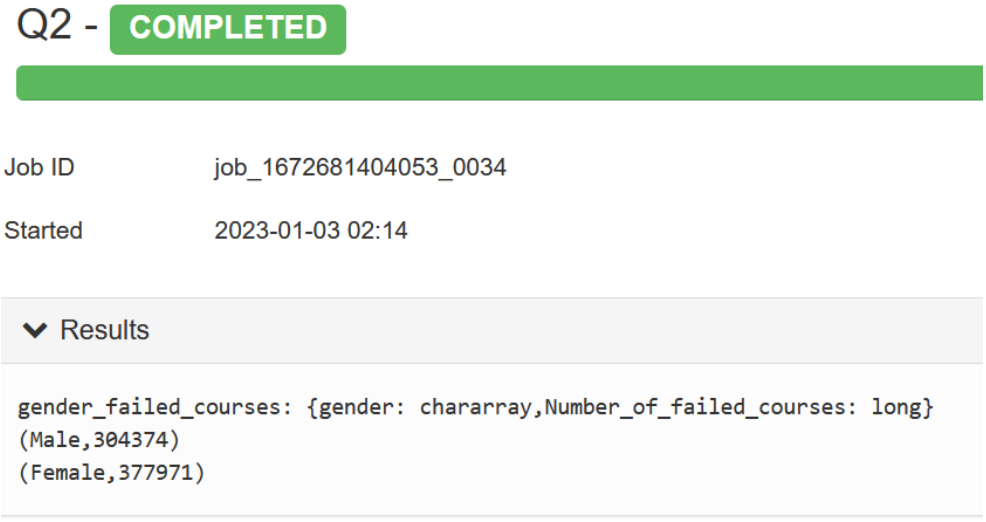
Two males, one from the public Mohammed Farouq secondary school and the other from the Egyptian English International School, received the highest scores ever in Egyptian schools.

Q2. What is the number of failed courses per gender?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Grouping the data.
grouped_data = GROUP Egyption_school BY gender;

--3. Get the number of failed courses per gender.
gender_failed_courses = FOREACH grouped_data GENERATE group AS gender,
COUNT(Egyption_school.no_of_failed_courses) AS Number_of_failed_courses;

--4. Describe the results.
DESCRIBE gender_failed_courses;

--5. Display the results.
DUMP gender_failed_courses;

--6. Store the results.
STORE gender_failed_courses INTO '/user/maria_dev/Q2_Result' USING
PigStorage (',');
```

- **The Result**

gender_failed_courses: {gender: chararray,Number_of_failed_courses: long}
(Male,304374)
(Female,377971)



- **Storing result**

- **Insight:**

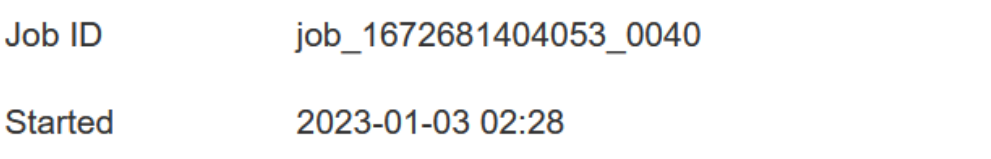
The number of failed courses for females is higher than the number of failed courses for males.

Q3. What is the number of people who can join the university based on their current major?

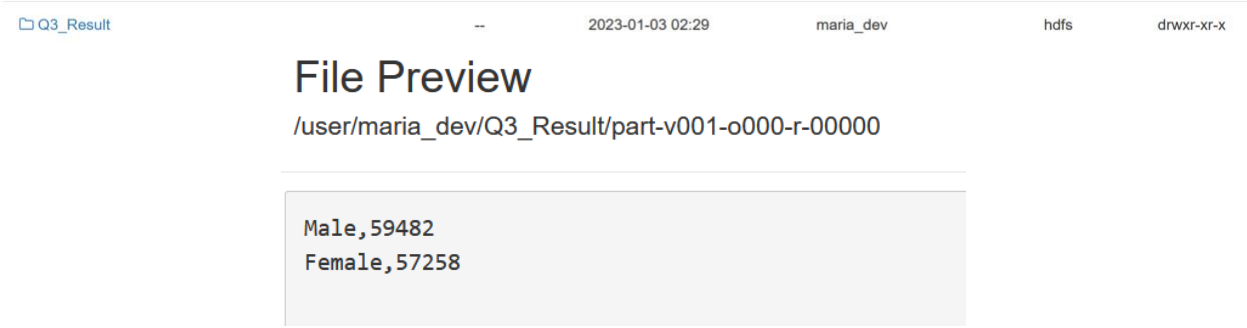
```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data based on joining condition.
filtered = FILTER Egyption_school BY (can_join_uni == 'Yes');

--3. Grouping the data.
grouped_data = GROUP filtered BY gender;

--4. Get the number of people who can join the university based on their
current major.
join_uni = FOREACH grouped_data GENERATE group AS gender,
COUNT(filtered.desk_no) AS gender_join_uni;

--5. Describe the results.
DESCRIBE join_uni;

--6. Display the results.
DUMP join_uni;

--7. Store the results.
STORE join_uni INTO '/user/maria_dev/Q3_Result' USING PigStorage (',');
```

- **The Result**

join_uni: {gender: chararray,gender_join_uni: long}
(Male,59482)
(Female,57258)



- **Storing result**



- **Insight:**

The number of males who can join university based on their major is higher than females.

Q4. What is the count of branches per city?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Grouping the data.
grouped_data = GROUP Egyption_school BY (city, branch);

--3. Get the count of students for each city per branch.
branch_count_data = FOREACH grouped_data GENERATE FLATTEN(group) AS (city,
branch), COUNT(Egyption_school.desk_no) AS branch_count;

--4. Order the result in descending order
order_by_data = ORDER branch_count_data BY branch_count DESC;

--5. Describe the results.
DESCRIBE order_by_data;

--6. Display the results.
DUMP order_by_data;

--7. Store the results.
STORE order_by_data INTO '/user/maria_dev/Q4_Result' USING PigStorage
(',');
```

- **The Result**

order_by_data: {city: chararray,branch: chararray,branch_count: long}
(Cairo,Literature,47165)
(Giza,Literature,38070)
(Cairo,Science - Health sciences,34247)
(Giza,Science - Health sciences,29574)
(Dakahlia,Science - Health sciences,26200)
(Sharqia,Science - Health sciences,25612)
(Beheira,Science - Health sciences,21719)
(Cairo,Science - Mathematics,21034)
(Alexandria,Literature,20468)
(Minya,Science - Health sciences,20330)
(Qalyubiyya,Literature,18973)
(Gharbia,Science - Health sciences,18544)
(Monufia,Science - Health sciences,18288)
(Alexandria,Science - Health sciences,17330)
(Sharqia,Literature,17080)
(Qalyubiyya,Science - Health sciences,16459)
(Minya,Literature,15492)
(Assiut,Science - Health sciences,14844)
(Dakahlia,Literature,14135)
(Kafr El Sheikh,Science - Health sciences,13817)
(Sohaj,Science - Health sciences,12993)
(Giza,Science - Mathematics,11954)
(Gharbia,Literature,11775)
(Monufia,Literature,10261)
(Faiyum,Science - Health sciences,9666)
(Beni Suef,Science - Health sciences,9559)
(Alexandria,Science - Mathematics,9421)
(Assiut,Literature,9203)
(Qena,Science - Health sciences,9131)
(Beheira,Literature,9081)
(Dakahlia,Science - Mathematics,7442)
(Qalyubiyya,Science - Mathematics,7302)
(Sharqia,Science - Mathematics,6804)
(Sohaj,Literature,6439)
(Kafr El Sheikh,Literature,5778)
(Beni Suef,Literature,5751)
(Qena,Literature,5574)
(Monufia,Science - Mathematics,5440)
(Damietta,Science - Health sciences,5271)
(Faiyum,Literature,5199)
(Gharbia,Science - Mathematics,5074)
(Ismailia,Science - Health sciences,4369)
(Aswan,Science - Health sciences,4343)
(Damietta,Literature,4130)
(Luxor,Science - Health sciences,4100)
(Ismailia,Literature,3469)
(Beni Suef,Science - Mathematics,2511)
(Minya,Science - Mathematics,2429)
(North Sinai,Science - Health sciences,2363)
(Beheira,Science - Mathematics,2329)
(Damietta,Science - Mathematics,2239)
(Luxor,Literature,2158)
(Aswan,Literature,2048)
(Assiut,Science - Mathematics,1900)
(Port Said,Literature,1832)
(Port Said,Science - Health sciences,1812)
(Kafr El Sheikh,Science - Mathematics,1677)
(Faiyum,Science - Mathematics,1595)
(Sohaj,Science - Mathematics,1558)
(Suez,Literature,1549)

(Ismailia,Science - Mathematics,1431)
(Suez,Science - Health sciences,1428)
(Qena,Science - Mathematics,1387)
(Matrouh,Science - Health sciences,1128)
(Red Sea,Literature,1108)
(Port Said,Science - Mathematics,1071)
(Red Sea,Science - Health sciences,1049)
(New Valley,Science - Health sciences,986)
(Suez,Science - Mathematics,953)
(Matrouh,Literature,914)
(New Valley,Literature,703)
(Aswan,Science - Mathematics,559)
(North Sinai,Literature,479)
(South Sinai,Science - Health sciences,436)
(Red Sea,Science - Mathematics,414)
(Luxor,Science - Mathematics,397)
(North Sinai,Science - Mathematics,328)
(South Sinai,Literature,311)
(Matrouh,Science - Mathematics,150)
(South Sinai,Science - Mathematics,108)
(New Valley,Science - Mathematics,95)

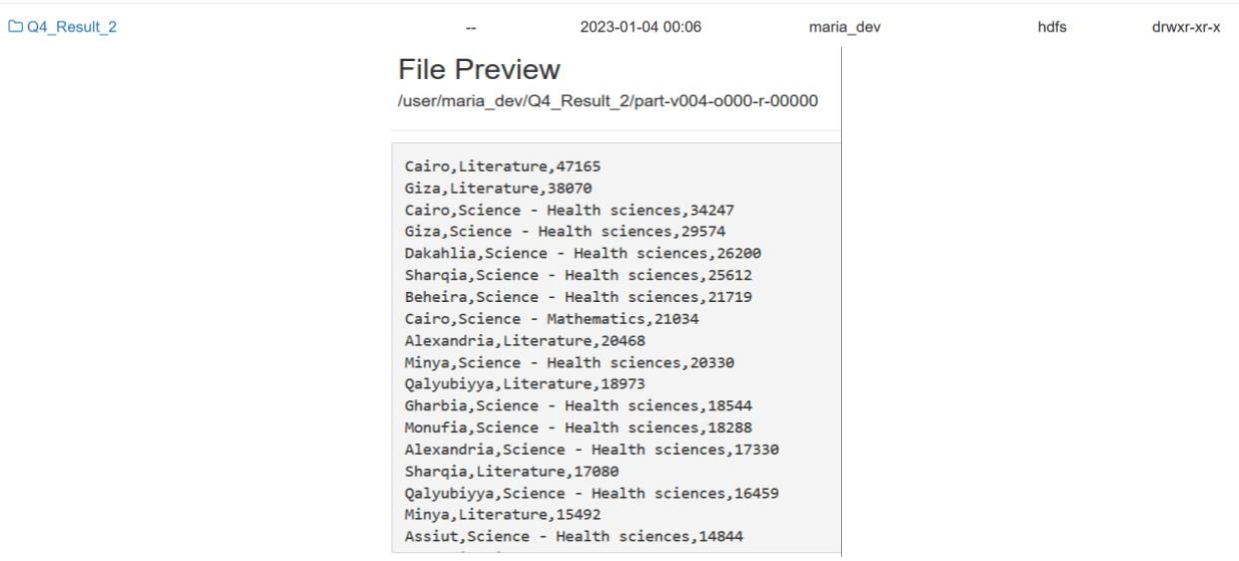
Q4 - COMPLETED

Job ID	job_1672681404053_0067
Started	2023-01-04 00:04

Results

```
order_by_data: {city: chararray,branch: chararray,branch_count: long}
(Cairo,Literature,47165)
(Giza,Literature,38070)
(Cairo,Science - Health sciences,34247)
(Giza,Science - Health sciences,29574)
(Dakahlia,Science - Health sciences,26200)
(Sharqia,Science - Health sciences,25612)
(Beheira,Science - Health sciences,21719)
(Cairo,Science - Mathematics,21034)
(Alexandria,Literature,20468)
(Minya,Science - Health sciences,20330)
(Qalyubiyya,Literature,18973)
(Gharbia,Science - Health sciences,18544)
(Monufia,Science - Health sciences,18288)
(Alexandria,Science - Health sciences,17330)
(Sharqia,Literature,17080)
(Qalyubiyya,Science - Health sciences,16459)
(Minya,Literature,15492)
(Assiut,Science - Health sciences,14844)
(Dakahlia,Literature,14135)
```

- **Storing result**



- **Insight:**

The highest number of students enrolled in all the branches exists in Cairo; apparently due to the population density in Cairo itself.

Q5. How many mixed schools per city?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
```

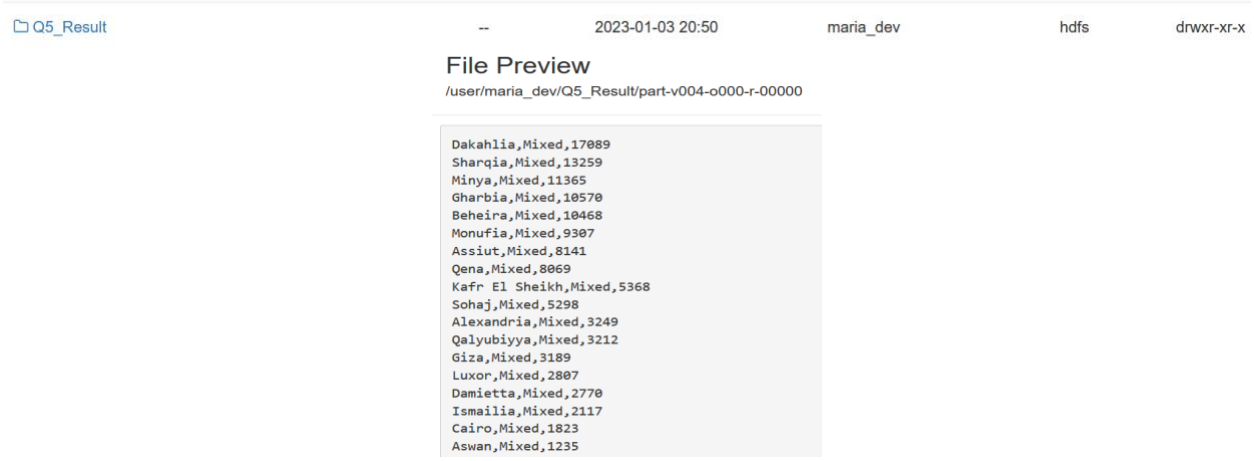
```
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data based on only the mixed schools.
filtered = FILTER Egyption_school BY (mixed_school=='Mixed');

--3. Grouping the data.
grouped_data = GROUP filtered BY (city, mixed_school);

--4. Get the count of students for each city per mixed school.
mixed_count_data = FOREACH grouped_data GENERATE FLATTEN(group) AS (city,
mixed_school), COUNT(filtered.desk_no) AS mixed_count;

--5. Order the result in descending order
order_by_data = ORDER mixed_count_data BY mixed_count DESC;

--6. Describe the results.
DESCRIBE order_by_data;

--7. Display the results.
DUMP order_by_data;

--8. Store the results.
STORE order_by_data INTO '/user/maria_dev/Q5_Result' USING PigStorage
(',');
```

- **The Result**

order_by_data: {city: chararray,mixed_school: chararray,mixed_count: long}
(Dakahlia,Mixed,17089)
(Sharqia,Mixed,13259)
(Minya,Mixed,11365)
(Gharbia,Mixed,10570)
(Beheira,Mixed,10468)
(Monufia,Mixed,9307)
(Assiut,Mixed,8141)
(Qena,Mixed,8069)
(Kafr El Sheikh,Mixed,5368)
(Sohaj,Mixed,5298)
(Alexandria,Mixed,3249)
(Qalyubiyya,Mixed,3212)
(Giza,Mixed,3189)
(Luxor,Mixed,2807)
(Damietta,Mixed,2770)
(Ismailia,Mixed,2117)
(Cairo,Mixed,1823)
(Aswan,Mixed,1235)
(Suez,Mixed,1106)
(New Valley,Mixed,606)
(Beni Suef,Mixed,392)
(North Sinai,Mixed,347)
(Red Sea,Mixed,345)
(Port Said,Mixed,301)

(Matrouh,Mixed,239)
(South Sinai,Mixed,29)

- **Storing result**



- **Insight:**

Dakahlia has the most mixed-gender schools, followed by Sharqia and Minya.

Q6. For students who did only the first attempt, how many of them failed?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data based on only the mixed schools.
filtered = FILTER Egyption_school BY (no_of_attempts=='First attempt
only') AND (status=='Failed');

--3. Grouping the data.
grouped_data = GROUP filtered BY (gender, no_of_attempts, status);

--4. Get the count of students for each city per mixed school.
failed_count_data = FOREACH grouped_data GENERATE FLATTEN(group) AS
(gender, no_of_attempts, status), COUNT(filtered.desk_no) AS failed_count;

--5. Order the result in descending order
order_by_data = ORDER failed_count_data BY failed_count DESC;

--6. Describe the results.
DESCRIBE order_by_data;

--7. Display the results.
DUMP order_by_data;

--8. Store the results.
STORE order_by_data INTO '/user/maria_dev/Q6_Result' USING PigStorage
(',');
```

- **The Result**

order_by_data: {gender: chararray,no_of_attempts: chararray,status: chararray,failed_count: long}
(Female,First attempt only,Failed,47672)
(Male,First attempt only,Failed,28583)

- **Storing result**

- **Insight:**

There were 28k male failures and 47k female failures among those who only made their first attempt.

Q7. How many 12 grade students have entered the public exams in the year 2022 from Egypt's capital city Cairo?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data based cairo city.
Filter_Cairo = FILTER Egyption_school BY (city == 'Cairo');

--3. Grouping the data.
student_group_all = GROUP Filter_Cairo ALL;

--4. Get the number of students who have entered the public exams in the
year 2022 from Egypt's capital city Cairo.
student_count = FOREACH student_group_all Generate
COUNT(Filter_Cairo.desk_no);

--5. Display the results.
DUMP student_count;

--7. Store the results.
STORE student_count INTO '/user/johara/pig_data/cairo_student_count' USING
PigStorage (',');
```

- **The Result**

There are 102,446 12 grade students at Cairo.

Q8. How many are the mixed, girls and boys schools in Egypt?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Grouping the data.
student_group_school_type = GROUP Egyption_school BY school_type;

--4. Get the number of the mixed, girls and boys schools in Egypt.
student_count = FOREACH student_group_school_type GENERATE group AS
School_type, COUNT(Egyption_school.desk_no)AS count_of_student;

--5. Display the results.
DUMP student_count;

--7. Store the results.
STORE student_count INTO '/user/johara/pig_data/school_type_student_count'
USING PigStorage (',');
```

- **The Result**

(Public School, 607660)

(Private School, 20268)

(International School, 54207)

(School for the Blind, 210)

Q9. What is the average score percentage for boys compared to girls?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where status is passed.
Filter_Passed = FILTER Egyption_school BY (status == 'Passed');

--3. Grouping the data.
student_group_gender = GROUP Filter_Passed by gender;

--4. Get the average score percentage for boys and girls.
student_average = FOREACH student_group_gender GENERATE group AS Gender,
AVG(Filter_Passed.percentage)AS average_persentage_by_gender;

--5. Display the results.
DUMP student_average;

--6. Store the results.
STORE student_average INTO '/user/johara/pig_data/student_average_gender'
USING PigStorage (',');
```

- **The Result**

(Male,71.86245011560318)

(Female,69.93731726966334)

Q10. What is the average score percentage by branch?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where status is passed.
Filter_Passed = FILTER Egyption_school BY (status == 'Passed');

--3. Grouping the data by the branch.
student_group_branch = GROUP Filter_Passed by branch;

--4. Get the average score percentage by each major.
student_average = FOREACH student_group_branch GENERATE group AS branch,
AVG(Filter_Passed.percentage)AS average_persentage_by_branch;

--5. Display the results.
DUMP student_average;

--6. Store the results.
STORE student_average INTO '/user/johara/pig_data/student_average_branch'
USING PigStorage (',');
```

- **The Result**

(Literature,65.86018891363379)

(Science - Mathematics,72.86361044358961)

(Science - Health sciences,73.59215815307552)

Q11. Which city has the highest average percentage rate? (Group by branch, calculate the percentage for each, sort from highest to lowest)

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where status is passed.
Filter_Passed = FILTER Egyption_school BY (status == 'Passed');

--3. Grouping the data by the city.
student_group_branch = GROUP Filter_Passed by city;

--4. Get the average score percentage for boys and girls.
student_average = FOREACH student_group_city GENERATE group AS branch, AVG
(Filter_Passed.percentage)AS average_persentage_by_city;

--5. Order the result in descending order.
order_student_average = ORDER student_average BY
average_persentage_by_city DESC;

--6. Display the results.
DUMP order_student_average;

--7. Store the results.
STORE order_student_average INTO
'/user/johara/pig_data/order_student_average' USING PigStorage (',');
```

- **The Result**

(North Sinai,77.82663808895063)

(Dakahlia,75.30098732616675)

(Port Said,74.35926065460868)

(Sohaj,73.53244228935543)

(Damietta,73.51924870761529)

(Kafr El Sheikh,72.93777186651181)

(Gharbia,72.85674858695481)

(South Sinai,72.75097558830184)

(Sharqia,72.34356036855999)

(Ismailia,71.58658994972612)

(Monufia,71.44064163723516)

(Beheira,70.99695253975155)

(New Valley,70.91270488285927)

(Qena,70.67285769376794)

(Matrouh,70.44108713695873)

(Faiyum,70.1241300538971)

(Assiut,69.38394651289266)

(Beni Suef,69.31378842472085)

(Cairo,69.3079294107257)

(Alexandria,69.29095425441005)

(Red Sea,69.21650197107445)

(Luxor,68.73408573507339)

(Qalyubiyya,68.64374760452269)

(Suez,68.35397799804123)

(Minya,67.96102100108463)

(Aswan,67.90657196553401)

(Giza,67.52480564162566)

Q12. What was the highest and lowest total points for girls compared to boys (filter the gender, filter1= boys, filter2= girls, get min and max total points for each filter)

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Grouping the data by gender.
group_gender = GROUP filtered BY gender;

--3. Get minimum and maximum percentages.
min_max_percentage = FOREACH group_gender GENERATE
MIN(Egyption_school.percentage) AS Min_percentage,
MAX(Egyption_school.percentage) AS Max_percentage;

--4. Describe the results.
DESCRIBE min_max_percentage;

--7. Display the results.
DUMP min_max_percentage;

--8. Store the results.
STORE min_max_percentage INTO '/user/johara/pig_data/min_max_percentage'
USING PigStorage (',');
```

- **The Result**

min_max_percentage: {Min_percentage: float,Max_percentage: float}
(0.0,99.27)
(0.0,99.02)

Q13. What is the number of homeschooled students?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where homeschooling is yes.
Filter_homeschooling = FILTER Egyption_school BY (homeschooling == 'Yes');

--3. Grouping the data.
student_group_all = GROUP Filter_homeschooling ALL;

--4. Get the number of homeschooled students.
student_count = FOREACH student_group_all  Generate
COUNT(Filter_homeschooling.desk_no);

--5. Display the results.
DUMP student_count;

--6. Store the results.
STORE student_count INTO
'/user/johara/pig_data/homeschooling_student_count' USING PigStorage
(',');
```

- **The Result**

(48428)

Q14. What is the lowest and highest score in Arabic subject?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Get the maximum score in the Arabic subject
max_arabic = FOREACH (GROUP Egyption_school ALL ) GENERATE
MAX(Egyption_school.arabic) as max_arabic;

--3. Get the minimum score in the Arabic subject
min_arabic = FOREACH (GROUP Egyption_school ALL ) GENERATE
MIN(Egyption_school.arabic) as min_arabic;

--4. Display the results.
DUMP max_arabic;
DUMP min_arabic;

--5. Store the results.
STORE max_arabic INTO '/user/Reema/Capestone/max_arabic.csv' count' USING
PigStorage (',');
STORE min_arabic INTO '/user/Reema/Capestone/min_arabic.csv' count' USING
PigStorage (',');
```

- **The Result**

(80.0)
(0.0)

Q15. What is the average score percentage for boys compared to girls with blindness?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where h where status is Passed and the school is for
the blind.
Filter_Passed_Blind = FILTER Egyption_school BY (status == 'Passed') and
(school_type == 'School for the Blind');

--3. Grouping the data by the gender.
student_group_gender = GROUP Filter_Passed_Blind BY gender;

--4. Get the average score percentage for boys compared to girls with
blindness.
student_average = FOREACH student_group_gender GENERATE group AS Gender,
AVG(Filter_Passed.percentage)AS average_persentage_by_gender;

--5. Display the results.
DUMP student_average;

--6. Store the results.
STORE student_average INTO '/user/Reema/Capestone/student_average.csv'
USING PigStorage (',');
```

- **The Result**

(Male,79.79792429366202)
(Female,84.24370789260007)

Q16. How many students failed in all of their courses?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where no_of_failed_courses is 10
filtered_data = FILTER Egyption_school BY (no_of_failed_courses == 10);

--3. Get the number of students who failed in all of their 10 courses.
failed_10_courses = FOREACH (GROUP filtered_data ALL) GENERATE
COUNT(filtered_data);

--4. Display the results.
DUMP failed_10_courses;

--5. Store the results.
STORE failed_10_courses INTO '/user/Reema/Capestone/failed_10_courses.csv'
USING PigStorage (',');
```

- **The Result**

(620)

Q17. How many schools for the blind per city?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where the school type is for blinds.
filtered_data = FILTER Egyption_school BY (school_type == 'School for the
Blind');

--3. Grouping the data by city and school type.
grouped_data = GROUP filtered_data BY (city, school_type);

--4. Get the number of schools for blind per city.
blind_count_data = FOREACH grouped_data GENERATE FLATTEN(group) AS (city,
school_type), COUNT(filtered.desk_no) AS blind_count;

--5. Order the result in descending order.
order_by_data = ORDER blind_count_data BY blind_count DESC;

--6. Describe the results.
DESCRIBE order_by_data;

--7. Display the results.
DUMP order_by_data;

--8. Store the results.
STORE order_by_data INTO '/user/Reema/Capestone/blind_school_per_city.csv'
USING PigStorage (',');
```

- **The Result**

order_by_data: {city: chararray,school_type: chararray,blind_count: long}
(Giza,School for the Blind,24)
(Minya,School for the Blind,19)
(Qalyubiyya,School for the Blind,16)
(Cairo,School for the Blind,15)
(Gharbia,School for the Blind,14)
(Faiyum,School for the Blind,13)
(Monufia,School for the Blind,12)
(Sohaj,School for the Blind,12)
(Alexandria,School for the Blind,12)
(Ismailia,School for the Blind,12)
(Sharqia,School for the Blind,8)
(Assiut,School for the Blind,7)
(Beheira,School for the Blind,7)
(Qena,School for the Blind,6)
(Dakahlia,School for the Blind,6)
(Kafr El Sheikh,School for the Blind,5)
(Suez,School for the Blind,5)
(Beni Suef,School for the Blind,4)
(Port Said,School for the Blind,4)
(Damietta,School for the Blind,4)
(Aswan,School for the Blind,4)
(Matrouh,School for the Blind,1)

Q18. How many Blind students who did only the first attempt, how many female and male who have failed?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data on only the first attempt, failed status, and blind
school
filtered_data = FILTER Egyption_school BY (no_of_attempts=='First attempt
only') AND (status=='Failed') AND (school_type == 'School for the Blind');

--3. Grouping the data.
grouped_data = GROUP filtered_data BY (gender, no_of_attempts,
status,school_type);

--4. Get the count of students for each city per Blind school.
failed_count_data = FOREACH grouped_data GENERATE FLATTEN(group) AS
(gender, no_of_attempts, status,school_type), COUNT (filtered.desk_no) AS
failed_count;

--5. Order the result in descending order.
order_by_data = ORDER failed_count_data BY failed_count DESC;

--6. Describe the results.
DESCRIBE order_by_data;

--7. Display the results.
```

```
DUMP order_by_data;

--8. Store the results.
STORE order_by_data INTO
'/user/Reema/Capestone/blind_students_who_failed.csv' USING PigStorage
(',');
```

- **The Result**

(Male,First attempt only,Failed,School for the Blind,2)

Q19. What is the average rate per branch for mixed school students?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where status is passed and mixed schools.
Filter_Passed = FILTER Egyption_school BY (status == 'Passed') and
(mixed_school=='Mixed');

--3. Grouping the data by the branch.
student_group_branch = GROUP Filter_Passed BY branch;

--4. Get the count of students for each city per Blind school.
student_average = FOREACH student_group_branch GENERATE group AS branch,
AVG(Filter_Passed.percentage)AS average_persentage_by_branch;

--5. Display the results.
DUMP student_average;

--8. Store the results.
STORE student_average INTO '/user/Reema/Capestone/mixed_student_average
.csv' USING PigStorage (',');
```

- **The Result**

(Literature,66.20555333606268)
(Science - Mathematics,73.80388388359036)
(Science - Health sciences,74.19587552280794)

Q20. How many students did failed in all their courses in mixed schools?

```
--1. Data Loading and Declaration (without the header).
Egyption_school = LOAD
'/user/maria_dev/capstone_project2_preprocessed_eng.csv'
 USING org.apache.pig.piggybank.storage.CSVExcelStorage(',',
'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
as (desk_no:int, gender:chararray, branch:chararray, city:chararray,
administration_translated:chararray,school_name_translated:chararray,
school_type:chararray, homeschooling:chararray, mixed_school:chararray,
no_of_attempts:chararray, no_of_failed_courses:int,can_join_uni:chararray,
final_grade:chararray, percentage:float,status:chararray, total:float,
arabic:float, first_foreign_lang:float, second_foreign_lang:float,
history:float,geography:float, philosophy:float, psychology:float,
biology:float, geology:float, chemistry:float, physics:float,
pure_mathematics:float, applied_math:float, religion:float,
national_education:float, economics_statistics:float,percentage_2nd:float,
status_2nd:chararray, total_2nd:float, arabic_2nd:float,
first_foreign_lang_2nd:float, second_foreign_lang_2nd:float,
history_2nd:float, geography_2nd:float, philosophy_2nd:float,
psychology_2nd:float,biology_2nd:float, geology_2nd:float,
chemistry_2nd:float, physics_2nd:float, pure_mathematics_2nd:float,
applied_math_2nd:float, religion_2nd:float, national_education_2nd:float,
economics_statistics_2nd:float);

--2. Filter the data where no_of_failed_courses is 10 and they are in
mixed school.
Filter_failed = FILTER Egyption_school BY (mixed_school=='Mixed') and
(no_of_failed_courses == 10);

--3. Get the count of students of mixed schools who failed on the 10
courses.
failed_10_courses_mixed = FOREACH (GROUP Filter_failed ALL) GENERATE
COUNT(Filter_failed);

--4. Display the results.
DUMP failed_10_courses_mixed;

--5. Store the results.
STORE failed_10_courses_mixed INTO
'/user/Reema/Capestone/mixed_student_average .csv' USING PigStorage (',');
```

- **The Result**