TASK REPORT
DATA SCIENCE AND ITS IMPLEMENTATION IN CODING

By:
Mohamad Irwan Afandi
Data Science Student

Data Fellowship
2020

Chapter 1
Introduction

Coding is one of the fundamentals that mast be mastered (at least know) by a data science. Why the data science need coding? The answer is, there are several task that we have to do like communication to server, get data from database, make data analysis, build a machine learning model, create an algorithm to solve problems, and soon. If we don't know about code, could we to be a professional data science? Think about it, so code is a must. As a beginner, we often ask "From the existing programming language, what must we learn for basic data science?" We can answer it by looking at the data science task. We know that the data science often work with database, algorithm and server. So the answer is SQL (for data base), Python (algorithm and build machine learning model) and the last is linux command + shell scripting (communicate with the server). Based on that problems, this report will show you some of real cases and how to solve it using code. But for the next time you need to explore it by yourself, at least you know what the code looks like. One think that you need to remember about coding, "***You don't need to memorize every lines of code but you need to practice as often as possible, after that the miracle will show you how easy it is. Keep learning guys!***"

Chapter 2
Progress Report

In this chapter you will have to fill in the table below according to the progress of the project that you have made along the way. We need to know how long it takes for you and how big the effort that you have done in order to complete this task. We appreciate detailed information.

Day/Date	**Task**	**Level** (easy/medium/hard)	**Comments**
11/08/2020	Doing all quizzes in iykra's website	easy	Takes a lot of time because too much subject need to learn
12/08/2020	Load and restore the dvd rental data into DBeaver, Solve the query problems	medium	Confused how to restore the data at the first time, takes time to understand the question and erd
13/08/2020	Python (create a function) and upload it into github.	Easy	Need more effort to design the read.me to make a better documentation on github.
14/08/2020	Create Summary and video for DB and Python, write in medium (how to install anaconda)	medium	New tools, write in medium :v
15/08/2020	Shell scripting, Create summary and video for shell scripting, write in github (basic linux command)	medium	Never learn linux command before, but it's fun
16/08/2020	Creating coding practice case report	easy	Feel so tired ☺

Chapter 3
Task Report

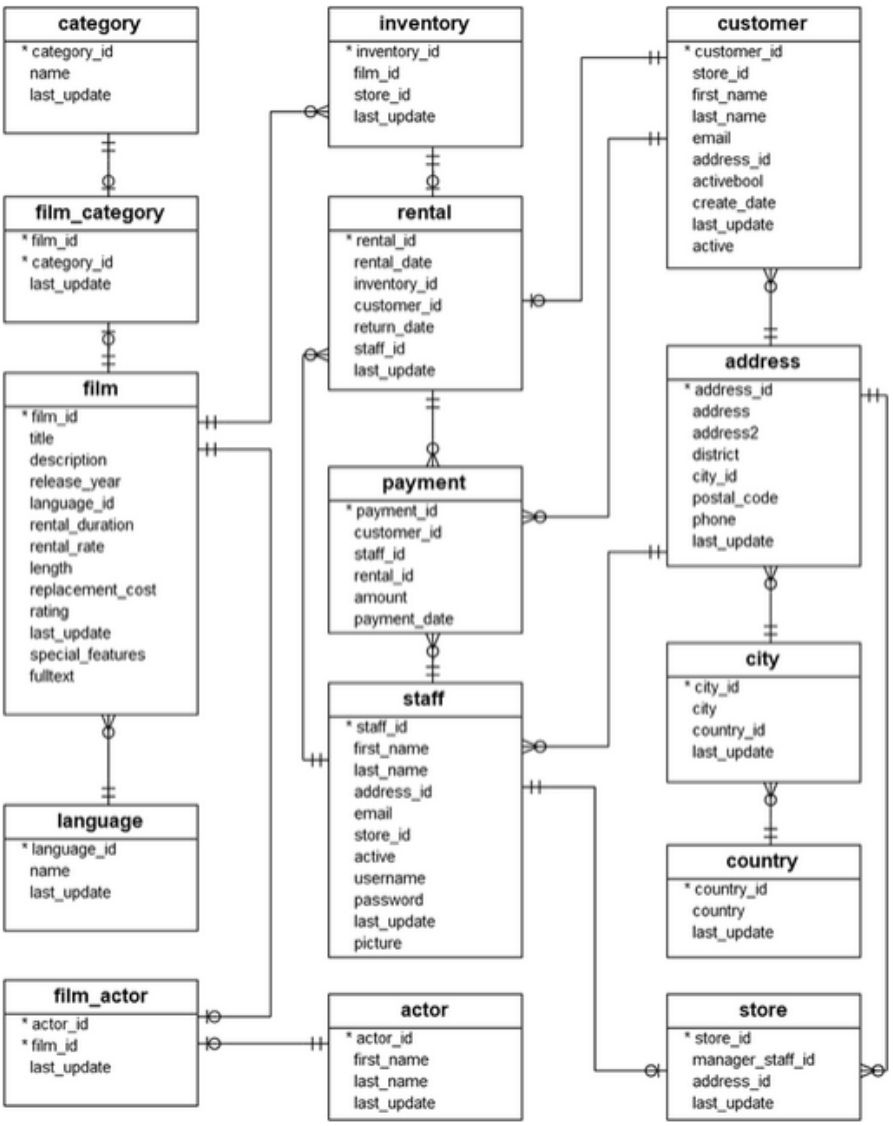
1. Please make SQL queries to answer these questions and post them on your Github account!

   Answer:

   The data used in this case study is DVD rental that can be downloaded on Postgresql website, or just access the link below.

   https://www.postgresqltutorial.com/postgresql-sample-database/

   The first, you need to understand the ER Model of the database to know the relations between tables. The following is the ER Model for the DVD rental Database.

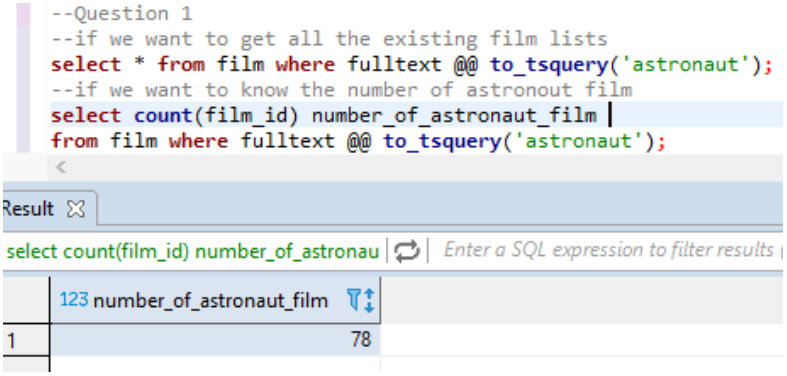


Furthermore, to answer the question we need to understand DML (Data Manipulation Language) syntax especially "select". In select syntax you also need to understand about where condition, aggregation function, grouping, having and ordering. The pattern from the select syntax can be seen below.
select col_name, aggregarion_function(col_name)
from tb_a join tb_b on tb_a.id = tb_b.a_id
where column_condition
group by col_name
having aggregation condition
order by col_name;

Now, we are ready to solve our task.

Question 1
A customer wants to know the films about "astronaut". How many recommendations could you give for him?

The purpose of this question is we need to count the number of films that the genre (from fulltext column) is astronaut. So we can find the data on film table with count of id_film as the aggregation function. The picture below shows the result of this question.

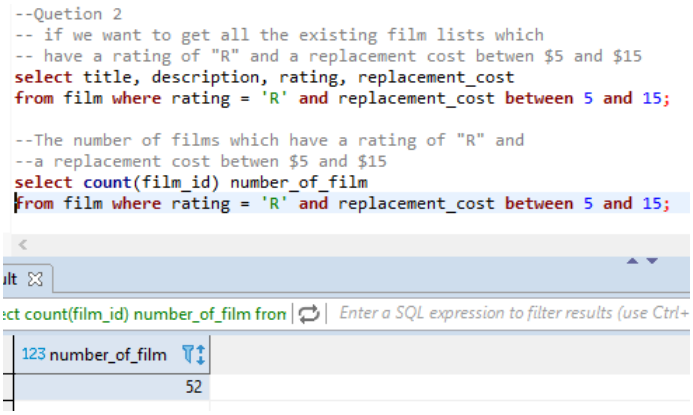


So the answer is, there are 78 films about Astronaut

Question 2
How many films have a rating of "R" and a replacement cost between $5 and 15?

The purpose of this question is we need to count the number of films which have "R" rating with condition the replacement cost between 5 and 15 (2 condition using and)

where we can get the data from table film. The picture below shows the result of this question.

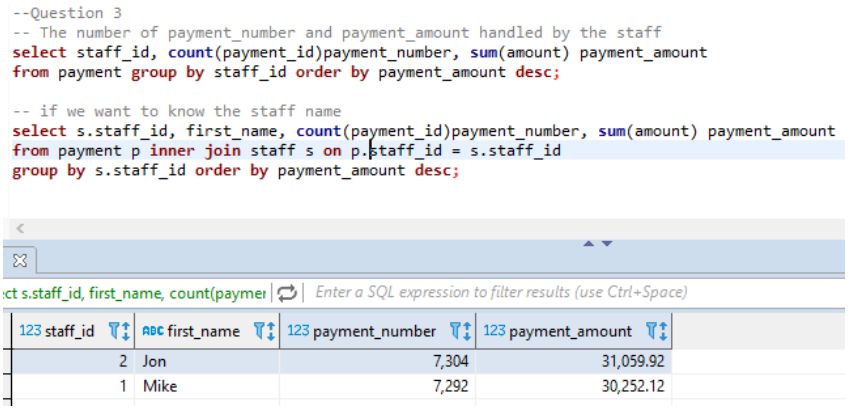


So the answer is, there are 52 films which have R rating and replacement cost between $5 and $15.

Question 3

We have two staff members with staff IDs 1 and 2. We want to give a bonus to the staff member that handled the most payments. How many payments did each staff member handle? And how much was the total amount processed by each staff member?

The purpose of this question is we need to calculate the payment transaction and the sum amount of that transaction group by staff id, and then short it descending. We can get this data from table payment join with table staff. The picture below shows the result of this question.



So we give the bonus to Jhon who can handle payment with the total amount $31.059,92

Question 4

Corporate headquarters is auditing the store, they want to know the average replacement cost of movies by rating.

The purpose of this question is we need to calculate the average replacement cost of movies group by rating and order the average ascending to make it easier to read. We can get the data from table film. The picture below shows the result of this question.

```
--Question 4
-- Find the average replacement cost of the movies by rating
select rating, avg(replacement_cost) average_cost
From film group by rating order by average_cost desc;
```

t rating, avg(replacement_cost) avera | Enter a SQL expression to filter results

rating	average_cost
PG-13	20.403
R	20.231
NC-17	20.138
G	20.125
PG	18.959

Question 5

We want to send coupons to the 5 customers who have spent the most amount of money. Get the customer name, email and their spent amount!

The purpose of this question is we need to get customer's name, email and sum of payment amount. Then short the sum of payment amount by descending and take first 5 customer. To get this data we need to join between table payment and customer. The picture below shows the result of this question.

```
--Question 5
--Find the customers name and email who have spent the most amount of money
select concat(c.first_name,' ',c.last_name) customer_name, email, sum(amount) total_payment_amount
from customer c inner join payment p on c.customer_id = p.customer_id
group by customer_name, email order by total_payment_amount desc limit 5;
```

t concat(c.first_name,' ',c.last_name) | Enter a SQL expression to filter results (use Ctrl+Space)

customer_name	email	total_payment_amount
Eleanor Hunt	eleanor.hunt@sakilacustomer.org	211.55
Karl Seal	karl.seal@sakilacustomer.org	208.58
Marion Snyder	marion.snyder@sakilacustomer.org	194.61
Rhonda Kennedy	rhonda.kennedy@sakilacustomer.org	191.62
Clara Shaw	clara.shaw@sakilacustomer.org	189.6

Now, we know that the coupons should send to Eleanor Hunt, Karl Seal, Marion Snyder, Rhonda Kennedy, and Clara Shaw.

Question 6

We want to audit our stock of films in all of our store. How many copies of each movie in each store do we have?

The purpose of this question is we need to count the number of film_id group by film_id and store_id that we get from table inventory and join with table film. We also can short the data by film_id by ascending or descending to make it easier to read. The picture below shows the result of this question.

```
--Question 6
--Get the copies number of each movies in each stores
select store_id, f.film_id, title, count(inventory_id) as copies_number
from inventory i inner join film f on i.film_id = f.film_id
group by  f.film_id, store_id order by f.film_id;
```

ntory(+)

ct store_id, f.film_id, title, count(inven | Enter a SQL expression to filter results (use Ctrl+Space

123 store_id	123 film_id	ABC title	123 copies_number
1	1	Academy Dinosaur	4
2	1	Academy Dinosaur	4
2	2	Ace Goldfinger	3
2	3	Adaptation Holes	4
1	4	Affair Prejudice	4
2	4	Affair Prejudice	3
2	5	African Egg	3
1	6	Agent Truman	3
2	6	Agent Truman	3
1	7	Airplane Sierra	2
2	7	Airplane Sierra	3
2	8	Airport Pollock	4
2	9	Alabama Devil	2
1	9	Alabama Devil	3

From this query, we get 1521 rows of data

Question 7

We want to know what customers are eligible for our platinum credit card. The requirements are that the customer has at least a total of 40 transaction payments. Get the customer name, email who are eligible for the credit card!

The purpose of this question is we need to get customer's name, email and the count of total transaction for every customer. Then filter the customers who have count of total

transaction at least 40 transaction. We can see that this case is aggregation with condition, so we use having to solve the aggregation's condition. We can get the data from table customer join with table payment. The picture below shows the result of this question.

```
--Question 7
--Find the customers name and email who have at least  40 transaction payment
select concat(c.first_name,' ',c.last_name) customer_name, email, count(payment_id) total_transaction
from customer c inner join payment p on c.customer_id = p.customer_id
group by customer_name, email having count(payment_id) >= 40  order by total_transaction desc;
```

omer

ct concat(c.first_name,' ',c.last_name) | Enter a SQL expression to filter results (use Ctrl+Space)

customer_name	email	total_transaction
Eleanor Hunt	eleanor.hunt@sakilacustomer.org	45
Karl Seal	karl.seal@sakilacustomer.org	42
Clara Shaw	clara.shaw@sakilacustomer.org	40

Now, we can give the platinum credit card to Eleanor Hunt, Karl Seal, and Clara Shaw.

Go to this link to check my query code :
https://github.com/irwanafandi24/IYKRA-Bootcamp
More specific:
https://github.com/irwanafandi24/IYKRA-Bootcamp/blob/master/Week%202%20(Coding%20SQL%20Python%20Shell%20Script)/dvd%20rental%20sql%20case.sql

2. Please make a python function that introduce your name, address, date of birth, and print them out into one sentence!

   Answer:

   To answer this question we need to understand how to build a function in python. So the pattern of the syntax is like this:

   def *funct_name (parm1, param2, paramn):*

   operation

   return *variable*

   And then to call the function is like this

   var_x = ***funct_name****(value1, value2, valuen)*

   This question asked to make input from keyboard based on the question which has been given. So we can use input method, note: input method always return string value.

   var_x = input("Question: ")

   This is the result from this task:

Create a function that can print out our data from input forms

```
In [1]: #define the function parameter
        def introduce(name, city, date):
            # %s, string formating
            return("My name is %s, I live in %s, I was born on %s" %(name, city, date))

In [2]: name = input("Input name: ")
        city = input("Input address: ")
        date = input("Input D.O.B: ")
        print("\nHai! "+introduce(name, city, date)) #calling the function

        Input name: Mohamad Irwan Afandi
        Input address: Banyuwangi, East Java
        Input D.O.B: November 24, 1996

        Hai! My name is Mohamad Irwan Afandi, I live in Banyuwangi, East Java, I was born on November 24, 1996
```

   Go to this link to check my query code :

   https://github.com/irwanafandi24/IYKRA-Bootcamp

   More specific:

   https://github.com/irwanafandi24/IYKRA-Bootcamp/blob/master/Week%202%20(Coding%20SQL%20Python%20Shell%20Script)/Fuction%20Example.ipynb

3. Write an article on Medium which explain python and jupyter notebook installation guide!
   Answer:
   This is the medium's link for jupyter notebook installation post:
   https://medium.com/@mohamadirwanafandi/start-playing-with-python-and-jupyter-notebook-on-windows-10-64-bit-8dea959c920e

4. Write an article on Medium which covers these subtopics:
   a. How to make a file and its content in linux
   b. Copy the file, then rename the second file
   c. Make a new directory, move the second file into the directory.
   d. Then finally remove the first directory.
   Answer:
   This is the medium's link for the basic linux commands post:
   https://medium.com/@mohamadirwanafandi/introduce-to-linux-commands-5529502c2cf3

Conclusion

There are a lot of way to go to Japan, no matter how. The most important thing is we can arrive in that country then take a selfie with tesla tower in Tokyo. It's same with the code, there are a lot of algorithm to solve a problem with the code and it depends on how we think. If there is an easier and more efficient way why do we choose an inefficient way? The question is how to get an effective one? *keep practice every day*.