**Database Final Project Report**

**Subject: Cellphone Rating System**

Member: 110652019林楷傑, 110550135黃孚翔, 110550152成文瀚,

110705013 沈昱宏, 110950014黃為碩

1.   An introduction of your application, including why you want to develop the application and the main functions of your application.

When we are choosing a cell phone, you have to compare their functions and their overall evaluation given by the other users. We develop a website for anyone who is having trouble choosing cell phones. In our application, you can have quick access to all the popular cell phone information, and you can also filter the cell phones according to your preferences. In addition, we have made some analysis based on a questionnaire, so you can know how the public or a particular group of people think of cell phones.

2.   Database design - describe the schema of all your tables in the database, including keys and index, if applicable (why you need the keys, or why you think that adding an index is or is not helpful).

一張含有 文字, 收據 的圖片

自動產生的描述

We use user\_id and cellphone\_id as keys in order to union the tables.

create table rate(

user\_id varchar(100),

    cellphone\_id int,

    rating real,

primary key (user\_id,cellphone\_id),

foreign key (user\_id) references users(user\_id),

foreign key (cellphone\_id) references data(cellphone\_id),

check(rating<=10 and rating >=0)

);

create table users(

user\_id varchar(100),

age real,

gender varchar(100),

occupation varchar(100),

password varchar(100),

primary key (user\_id),

check(gender in ('Male','Female','null'))

);

create table data(

cellphone\_id int,

brand varchar(100),

model varchar(100),

internal\_memory real,

RAM real,

performance real,

main\_camera real,

selfie\_camera real,

battery\_size real,

screen\_size real,

weight real,

price real,

release\_date date,

primary key(cellphone\_id)

);

3.   Database design - describe the normal form of all your tables. If the tables are not in BCNF, please include the reason for it (performance trade-off, etc.).

After we normalize the “data” table, our tables are all in BCNF, since all redundancy based on functional dependency has been removed.

The function dependency in users:  
(user\_id)→(all) and user-id is the superkey of users ,so users is in BCNF.

The function dependency in rate:

(user\_id,cellphon\_id)→(all) similarly rate is in BCNF.

The function dependency in data:

   (cellphone\_id)→(all)   similarly data is in BCNF.

4.   From the data sources to the database - describe the data source and the original format.

Here is the data source:

<https://www.kaggle.com/datasets/meirnizri/cellphones-recommendations?select=cellphones+users.csv>

The data source consists of three csv. files:

1. cellphone users (user\_id, age, gender, occupation)
2. cellphone ratings (user\_id, cellphone\_id, rating)
3. cellphone data (cellphone\_id, brand, model, operating system, internal memory, RAM, performance, main camera, selfie camera, battery size, screen size, weight, price, release date)

In cellphone data, there is originally one attribute for the operating system and a simple function dependency (brand→operating system, apple uses ios, others use android). Therefore, it violates BCNF, but we won’t use the operating system, so operating is not in our data table.

5.   From the data sources to the database - describe the methods of importing the original data to your database and strategies for updating the data, if you have one.

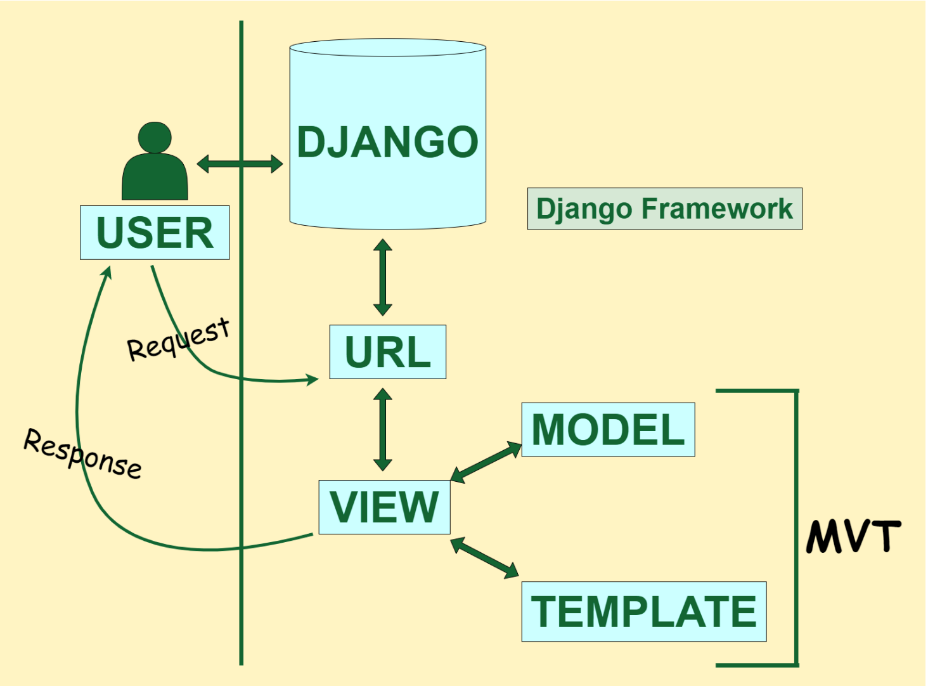
We use Amazon RDS to maintain our database, then use PGAdmin to connect this database. We first try to import our original data through PGAdmin. Since there are several errors in the previous data, we use pandas in python to deal with all the errors (all the details are in the python\_modified file). The resource of our database will never be updated, so we decided to update all the data by ourselves or the other users. We can either use PGAdmin or the application to update the data. In our application, we will have a page for data updating. You can input the data according to the format, and the data will be sent into the database.

6.   Application with database - explain why your application needs a database.

A database can help our application store and retrieve data in an organized and structured way. This can make it easier for our application to access and manipulate the data as needed. Without a database, our application can’t store a large amount of data and manage the data efficiently, which might result in difficulties in analyzing the cell phones.

7.   Application with database - includes the queries that are performed by your application, how your application performed these queries (connections between application and database), and what are the cooperating functions for your application.

* Backend:

We built this website with django as backend architecture and connected postgresql with django such that we can fetch the data from the database in the backend. We did not deploy this application to the server since we did not have time to study how to do it, so it can be seen only on the local end. But we will deploy the application to the AWS ec2 in the future.

* Queries and cooperating functions:  
  We have implemented some functions for users to become familiar with data and check their demand. For example, finding the highest rating cellphone. We implemented the SQL query and ran it in the backend using the Python package psycopg2. Here we list all the query we have used:

A. cellphone average rating (with average of all ratings having

cellphone\_id -1)

(select 'average' as model,round(cast(avg(rating) as decimal),3) as average

from rate,data)

union

(select model,newt.average

from(select cellphone\_id,round(cast(avg(rating) as decimal),3) as average

from rate

group by cellphone\_id) as newt,data

where newt.cellphone\_id=data.cellphone\_id)

order by average desc;

1. favorite cell phone of users

(count the cell which has the highest rating given by individual user)

select model,newbigt.number\_of\_users

from data,(select cellphone\_id,count(user\_id) as number\_of\_users

from (select rate.user\_id, rate.cellphone\_id

from (select user\_id,max(rating) as highest\_rating

from rate

group by user\_id) as highest\_rate, rate

where highest\_rate.user\_id=rate.user\_id and rate.rating=highest\_rate.highest\_rating) as newT

group by newT.cellphone\_id)as newbigt

where newbigt.cellphone\_id = data.cellphone\_id

order by number\_of\_users DESC

1. amount of cellphone owned by all users

select model,newt.count

from(select cellphone\_id,count(user\_id)

from rate

group by cellphone\_id)as newt,data

where data.cellphone\_id=newt.cellphone\_id

order by newt.count desc;

1. list the cell phones which has average rating bigger than total average

select model,newt.average

from(select cellphone\_id as cellphone\_id,cellphones.average

from(select cellphone\_id,round(cast(avg(rating) as decimal),3) as average

from rate

group by cellphone\_id) as cellphones, (select round(cast(avg(rating) as decimal),3) as average from rate) as tavg

where cellphones.average>tavg.average)as newt,data

where newt.cellphone\_id=data.cellphone\_id

order by newt.average desc;

1. market share -brand (in this survey)

with brand\_amount(brand,amount) as

(select data.brand, count(\*) as amount

from data

join rate on data.cellphone\_id=rate.cellphone\_id

group by data.brand

order by amount desc)

select brand\_amount.brand,

concat(round(cast(brand\_amount.amount as decimal)/9.9,2),'%') as market\_share

from brand\_amount

order by brand\_amount.amount desc

1. market share- operating system

with op(operating\_system,amount) as

(select data.operating\_system, count(\*) as amount

from data

join rate on data.cellphone\_id=rate.cellphone\_id

group by data.operating\_system

order by amount desc)

select op.operating\_system,

concat(round(cast(op.amount as decimal)/9.9,3),'%') as market\_share\_operating\_system

from op

order by op.amount desc

1. list the average rating & number of ratings given by females

select data.model,tablelast.average\_rating,tablelast.number\_of\_ratings

from(select cellphone\_id,round(cast(avg(rating) as decimal),3) as average\_rating,count(user\_id) as number\_of\_ratings

from (select \* from(select \* from rate NATURAL join users) as bigT where gender='Female')as newbigT

group by cellphone\_id)as tablelast,data

where tablelast.cellphone\_id=data.cellphone\_id

order by tablelast.average\_rating desc;

1. list the top 10 cellphone for a particular occupation

select data.model,newt.average

from(select round(cast(avg(rating) as decimal),3) as average,cellphone\_id

from(select \* from rate NATURAL join users where occupation='information technology') as bigt

group by cellphone\_id

limit 10)  as newt,data

where data.cellphone\_id=newt.cellphone\_id

order by newt.average desc;

1. list the top 10 cellphone for elders (with age>average age)

select data.model,newt.averageRate

from(select cellphone\_id,round(cast(avg(rating) as decimal),3) as averageRate

from(select \* from rate natural join (select user\_id from

(select avg(age) as avg\_age from users) as avgT,users

where users.age>avgT.avg\_age) as newt) as bigT

group by cellphone\_id

limit 10) as newt,data

where newt.cellphone\_id=data.cellphone\_id

order by newt.averageRate desc;

* Frontend is built by HTML, Javascript and CSS. We did our best to make the user experience good.

8.   All the other details of your application that you want us to know.

To update the data, we create some pages, like login, register, and rating page.

To access the operation page in our application, users need to login. In addition, our users table needs to have a record about this user. If not, the user needs to register, which will update the users table. Then after login, users can rate any cell phone they want to rate in the rating page. If the data table has the record about this table, rate will be updated, otherwise, data will be updated and contain the new record for some new cell phones.