**TEAM MEMBERS**

**Plagiarism Policy**

Plagiarism is the representation of another's words, ideas, or work as one’s own. It is the student’s responsibility to self-educate on what constitutes plagiarism. All students are responsible for reading and understanding these rules and abide by the signed and submitted Academic Integrity Statement.

* Students may not show their solutions to the homework assignments to other students before the submission deadline.
* Students may discuss lecture materials, but they may not discuss homework assignments with their classmates.
* Unless otherwise stated, students may collaborate on assignments with others within the same group. For this course, all homework assignments are group efforts and only the project is a joint effort venture.

All submissions should be the original works. Any incidence of plagiarism (any exchange, reuse, or direct submission of other’s work, e.g., documents, code etc., as your own or as your team’s) or other academic dishonesty will be discussed with the student. Students found guilty of plagiarism can expect a Drexel University Alleged Academic Misconduct Report to be filed to the University Judicial Office, where a permanent record is maintained and result in an F for this course.

**Academic Integrity Statement**

To ensure adherence to the academic honesty policies, all students are required to sign and submit the following statement.

I certify that:

* Homework, project, exams and/or all other assignments in this course will be entirely my own original work, produced in response to the specific course requirements.
* I will not quote the words of any other person from a printed source or a website without indicating what has been quoted and providing an appropriate citation.
* I will not submit my work in this course to satisfy the requirements of any other course nor will I use the same work from any other course to satisfy the requirements of this course.

|  |  |  |  |
| --- | --- | --- | --- |
| **Full Name** | **Username (abc123)** | **Date** | **Signature (initials)** |
| Anamika Rekha | ar3876 | 03/05/2023 | AR |
| Abhishekh Mohan Patil | ap3869 | 03/05/2023 | AP |
| Vijval Vemula | vv354 | 03/05/2023 | VV |
| Priyanka Patil | pp673 | 03/05/2023 | PP |
|  |  |  |  |

Obesity Data

**Overall instructions**

Consider the data model “Obesity Data” shared on Vertabelo.com:

* + **OBESITY\_DATA** table is the raw data imported directly from the available CSV file.
  + All other tables represent the normalized version of the same data.

Diagram

Description automatically generated

The response for each of the following questions must be given by two queries:

* One using the raw data (**hc848.OBESITY\_DATA** table only)
  + You **MUST** use prefix “**hc848**” in the first query
* Another query (only) using tables from the normalized data model
  + You **CAN’T** use **OBESITY\_DATA**
  + You **CAN’T** use prefix “**hc848”**

**DO NOT use PK and FK columns as filters (WHERE), as I will be testing your queries in a different database that uses different ID values.**

Each student MUST work in one question in each section below, and then agree with your group regarding the final answer to be submitted. If your group has only 4 members, you can submit only 4 questions in each section. Likewise, if your group has only 2 or 3 members.

# Section 1

|  |  |
| --- | --- |
| Question 1 | **How many males are above 30 years old and eat high calorie food?** |
| Student | Vv354 |
| SQL 1 | select count(\*) as required\_count\_raw  from obesity\_data  where gender = 'Male'  and age > 30  and eat\_high\_caloric\_food = 'yes'; |
| SQL 2 | select count(\*) as required\_count\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  join obesity\_boolean ob on op.eat\_high\_caloric\_food\_id = ob.boolean\_id  where og.gender\_name like '%Male%'  and op.age > 30  and ob.boolean\_value like '%Yes%'; |
| Group Submission | |
| SQL 1 | select count(\*) as required\_count\_raw  from obesity\_data  where gender = 'Male'  and age > 30  and eat\_high\_caloric\_food like '%yes%'; |
| Screenshot |  |
| SQL 2 | select count(\*) as required\_count\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  join obesity\_boolean ob on op.eat\_high\_caloric\_food\_id = ob.boolean\_id  where og.gender\_name like '%Male%'  and op.age > 30  and ob.boolean\_value like '%Yes%'; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 2 | **How many people smoke and drink frequently and are overweighted?** |
| Student | Ar3876 |
| SQL 1 | select count(\*) as required\_count\_raw  from obesity\_data  where smoke like '%yes%'  and alcohol\_frequency like '%Frequently%'  and weight\_status like '%Overweight%'; |
| SQL 2 | select count(\*) as required\_count\_norm  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_boolean ob on op.smoke\_id = ob.boolean\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  where ob.boolean\_value like '%Yes%'  and oof.frequency\_description like '%Frequently%'  and ows.weight\_status\_description like '%Overweight%'; |
| Group Submission | |
| SQL 1 | select count(\*) as required\_count\_raw  from obesity\_data  where smoke like '%yes%'  and alcohol\_frequency like '%Frequently%'  and weight\_status like '%Overweight%'; |
| Screenshot |  |
| SQL 2 | select count(\*) as required\_count\_norm  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_boolean ob on op.smoke\_id = ob.boolean\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  where ob.boolean\_value like '%Yes%'  and oof.frequency\_description like '%Frequently%'  and ows.weight\_status\_description like '%Overweight%'; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 3 | **What is the percentage of patient who smoke?** |
| Student | ap3869 |
| SQL 1 | select count(\*) as total\_patients\_raw,  round(100 \* (sum(case when smoke like '%yes%' then 1 else 0 end) / count(\*)), 2) as smoking\_percentage\_raw  from obesity\_data; |
| SQL 2 | select count(\*) as total\_patients\_norm,  round(100 \* (sum(case when ob.boolean\_value like '%Yes%' then 1 else 0 end) / count(\*)), 2) as smoking\_percentage\_norm  from obesity\_patient op  join obesity\_boolean ob on op.smoke\_id = ob.boolean\_id; |
| Group Submission | |
| SQL 1 | select count(\*) as total\_patients\_raw,  round(100 \* (sum(case when smoke like '%yes%' then 1 else 0 end) / count(\*)), 2) as smoking\_percentage\_raw  from obesity\_data; |
| Screenshot |  |
| SQL 2 | select count(\*) as total\_patients\_norm,  round(100 \* (sum(case when ob.boolean\_value like '%Yes%' then 1 else 0 end) / count(\*)), 2) as smoking\_percentage\_norm  from obesity\_patient op  join obesity\_boolean ob on op.smoke\_id = ob.boolean\_id; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 4 | **What is the percentage of patients that do monitor their daily calories intake and work out in each gender?** |
| Student | ar3876 |
| SQL 1 | select count(\*) as total,gender,    round(100 \* (sum(case when monitor\_calories\_daily like '%yes%' and avg\_days\_activity\_weekly >0 then 1 else 0 end) / count(\*)), 2) as per    from hc848.OBESITY\_DATA group by gender; |
| SQL 2 | select count(\*) as total\_patients\_norm,og.gender\_name,  round  (  100 \* (sum(  case when ob.boolean\_value like '%Yes%'  and op.avg\_days\_activity\_weekly > 0 then 1 else 0 end  ) / count(\*))  , 2) as percent  from obesity\_patient op  join obesity\_boolean ob on op.monitor\_calories\_daily\_id = ob.boolean\_id  join obesity\_gender og on og.gender\_id=op.gender\_id  group by og.gender\_name; |
| Group Submission | |
| SQL 1 | select count(\*) as total\_patients\_raw,gender,  round  (  100 \* (sum(  case when monitor\_calories\_daily like '%yes%'  and avg\_days\_activity\_weekly > 0 then 1 else 0 end  ) / count(\*))  , 2) as percent  from hc848.obesity\_data  group by gender; |
| Screenshot |  |
| SQL 2 | select count(\*) as total\_patients\_norm,og.gender\_name,  round  (  100 \* (sum(  case when ob.boolean\_value like '%Yes%'  and op.avg\_days\_activity\_weekly > 0 then 1 else 0 end  ) / count(\*))  , 2) as percent  from obesity\_patient op  join obesity\_boolean ob on op.monitor\_calories\_daily\_id = ob.boolean\_id  join obesity\_gender og on og.gender\_id=op.gender\_id  group by og.gender\_name; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 5 | **What is the average height and weight for each of the gender?** |
| Student | pp673 |
| SQL 1 | select count(\*) as total,gender,round(avg(height\_m),2) as Height\_Average,round(avg(weight\_kg),2) as Weight\_Average  from hc848.OBESITY\_DATA group by gender; |
| SQL 2 | SELECT count(\*) as total,OG.GENDER\_ID,OG.GENDER\_NAME,round(avg(OP.height\_m),2) as Height\_Average,round(avg(OP.weight\_kg),2)  from OBESITY\_PATIENT OP  JOIN OBESITY\_GENDER OG ON OG.GENDER\_ID=OP.GENDER\_ID  group by OG.GENDER\_ID,OG.GENDER\_NAME; |
| Group Submission | |
| SQL 1 | select gender,  round(avg(height\_m), 2) as avg\_height\_m\_raw,  round(avg(weight\_kg), 2) as avg\_weight\_kg\_raw  from HC848.obesity\_data  group by gender; |
| Screenshot |  |
| SQL 2 | select og.gender\_name as gender,  round(avg(op.height\_m), 2) as avg\_height\_m\_norm,  round(avg(op.weight\_kg), 2) as avg\_weight\_kg\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  group by og.gender\_name; |
| Screenshot |  |

# Section 2

|  |  |
| --- | --- |
| Question 6 | **How many main meals daily do patients who use automobiles for transportation have in average?** |
| Student | ap3869 |
| SQL 1 | SELECT ROUND(avg(MAIN\_MEALS\_DAILY),2) AS MAINMEALSDAILYAUTOPERC  FROM hc848.OBESITY\_DATA  WHERE TRANSPORTATION\_METHOD LIKE '%Automobile%'; |
| SQL 2 | SELECT ROUND(AVG(OP.MAIN\_MEALS\_DAILY),2) AS MAINMEALSDAILYAUTOPERC  FROM OBESITY\_PATIENT OP  JOIN OBESITY\_TRANSP\_METHOD OTM ON OTM.TRANSP\_METHOD\_ID=OP.TRANSP\_METHOD\_ID  WHERE OTM.TRANSP\_METHOD\_NAME LIKE '%Automobile%'; |
| Group Submission | |
| SQL 1 | select transportation\_method,  round(avg(main\_meals\_daily), 2) as avg\_meals\_daily\_raw  from hc848.obesity\_data  group by transportation\_method  order by transportation\_method  fetch first 1 rows only; |
| Screenshot |  |
| SQL 2 | select otm.transp\_method\_name as transportation\_method,  round(avg(op.main\_meals\_daily), 2) as avg\_meals\_daily\_norm  from obesity\_patient op  join obesity\_transp\_method otm on op.transp\_method\_id = otm.transp\_method\_id  group by otm.transp\_method\_name  order by otm.transp\_method\_name  fetch first 1 row only; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 7 | **For each of the gender, based on their smoking and drinking preference, how many people are overweight or obese?** |
| Student | pp673 |
| SQL 1 | SELECT GENDER, SMOKE, ALCOHOL\_FREQUENCY,  COUNT(\*) AS COUNT\_OVERWEIGHT\_OR\_OBESE  FROM OBESITY\_DATA  WHERE WEIGHT\_STATUS LIKE '%Overweight%'   OR WEIGHT\_STATUS LIKE '%Obesity%'  GROUP BY GENDER, SMOKE, ALCOHOL\_FREQUENCY  ORDER BY COUNT\_OVERWEIGHT\_OR\_OBESE DESC; |
| SQL 2 | SELECT g.GENDER\_NAME, b.BOOLEAN\_VALUE AS SMOKE, f.FREQUENCY\_DESCRIPTION AS ALCOHOL\_FREQUENCY,  COUNT(\*) AS COUNT\_OVERWEIGHT\_OR\_OBESE  FROM hc848.OBESITY\_PATIENT p  JOIN hc848.OBESITY\_GENDER g ON p.GENDER\_ID = g.GENDER\_ID  JOIN hc848.OBESITY\_BOOLEAN b ON p.SMOKE\_ID = b.BOOLEAN\_ID  JOIN hc848.OBESITY\_FREQUENCY f ON p.ALCOHOL\_FREQUENCY\_ID = f.FREQUENCY\_ID  JOIN hc848.OBESITY\_WEIGHT\_STATUS w ON p.WEIGHT\_STATUS\_ID = w.WEIGHT\_STATUS\_ID  WHERE w.WEIGHT\_STATUS\_DESCRIPTION LIKE '%Overweight%' OR w.WEIGHT\_STATUS\_DESCRIPTION LIKE '%Obesity%'  GROUP BY g.GENDER\_NAME, b.BOOLEAN\_VALUE, f.FREQUENCY\_DESCRIPTION  ORDER BY COUNT\_OVERWEIGHT\_OR\_OBESE DESC; |
| Group Submission | |
| SQL 1 | select gender, smoke, alcohol\_frequency,  count(\*) as total\_count\_raw,  sum(case when weight\_status like '%Overweight%'  or weight\_status like '%Obesity%' then 1 else 0 end) as overweight\_and\_obese\_raw,  round(100 \* (sum(case when weight\_status like '%Overweight%'  or weight\_status like '%Obesity%' then 1 else 0 end) / count(\*)), 2) as percent\_overweight\_and\_obese\_raw  from hc848.obesity\_data  group by gender, smoke, alcohol\_frequency  order by overweight\_and\_obese\_raw desc; |
| Screenshot |  |
| SQL 2 | select og.gender\_name as gender,  ob.boolean\_value as smoke,  oof.frequency\_description as alcohol\_frequency,  count(\*) as total\_count\_norm,  sum(case when ows.weight\_status\_description like '%Overweight%'  or ows.weight\_status\_description like '%Obesity%' then 1 else 0 end) as overweight\_and\_obese\_norm,  round(100 \* (sum(case when ows.weight\_status\_description like '%Overweight%'  or ows.weight\_status\_description like '%Obesity%' then 1 else 0 end) / count(\*)), 2) as percent\_overweight\_and\_obese\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  join obesity\_boolean ob on op.smoke\_id = ob.boolean\_id  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  group by og.gender\_name, ob.boolean\_value, oof.frequency\_description  order by overweight\_and\_obese\_norm desc; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 8 | **Rank the frequency of alcohol consumption for the obese people (largest to smallest).** |
| Student | vv354 |
| SQL 1 | select alcohol\_frequency,  count(\*) as alcohol\_freq\_count\_raw,  rank() over(order by count(\*) desc) as rank\_raw  from obesity\_data  where weight\_status like '%Obesity%'  group by alcohol\_frequency; |
| SQL 2 | select case  when oof.frequency\_description like '%Always%' then 'Always'  when oof.frequency\_description like '%Frequently%' then 'Frequently'  when oof.frequency\_description like '%Sometimes%' then 'Sometimes'  else 'no'  end as alcohol\_frequency,  count(\*) as alcohol\_freq\_count\_norm,  rank () over(order by count(\*) desc) as rank\_norm  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  where ows.weight\_status\_description like '%Obesity%'  group by oof.frequency\_description; |
| Group Submission | |
| SQL 1 | select alcohol\_frequency,  count(\*) as alcohol\_freq\_count\_raw,  rank() over(order by count(\*) desc) as rank\_raw  from obesity\_data  where weight\_status like '%Obesity%'  group by alcohol\_frequency; |
| Screenshot |  |
| SQL 2 | select case  when oof.frequency\_description like '%Always%' then 'Always'  when oof.frequency\_description like '%Frequently%' then 'Frequently'  when oof.frequency\_description like '%Sometimes%' then 'Sometimes'  else 'no'  end as alcohol\_frequency,  count(\*) as alcohol\_freq\_count\_norm,  rank () over(order by count(\*) desc) as rank\_norm  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  where ows.weight\_status\_description like '%Obesity%'  group by oof.frequency\_description; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 9 | **What is the average hours spent on using tech devices for individuals in their 20s based on their gender?** |
| Student | ar3876 |
| SQL 1 | select age,  round(avg(case when gender like '%Male%' then avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_male\_raw,  round(avg(case when gender like '%Female%' then avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_female\_raw  from hc848.obesity\_data  where age between 20 and 29  group by age  order by age desc; |
| SQL 2 | select op.age,  round(avg(case when og.gender\_name like '%Male%' then op.avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_male\_norm,  round(avg(case when og.gender\_name like '%Female%' then op.avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_female\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  where op.age between 20 and 29  group by op.age  order by op.age desc; |
| Group Submission | |
| SQL 1 | select age,  round(avg(case when gender like '%Male%' then avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_male\_raw,  round(avg(case when gender like '%Female%' then avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_female\_raw  from hc848.obesity\_data  where age >= 20 and age <= 29  group by age  order by age desc; |
| Screenshot |  |
| SQL 2 | select op.age,  round(avg(case when og.gender\_name like '%Male%' then op.avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_male\_norm,  round(avg(case when og.gender\_name like '%Female%' then op.avg\_hours\_tech\_device\_daily else null end), 2) as avg\_tech\_hours\_female\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  where op.age >= 20 and op.age <= 29  group by op.age  order by op.age desc; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 10 | **For each of the weight status, what is the average days spent on physical activities for each gender? Use different columns to provide gender information.** |
| Student | vv354 |
| SQL 1 | <student initial proposal> select weight\_status,  round(avg(case when gender like '%Male%' then avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_male\_raw,  round(avg(case when gender like '%Female%' then avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_female\_raw  from obesity\_data  group by weight\_status  order by weight\_status; |
| SQL 2 | select ows.weight\_status\_description as weight\_status,  round(avg(case when og.gender\_name like '%Male%' then op.avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_male\_norm,  round(avg(case when og.gender\_name like '%Female%' then op.avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_female\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  group by ows.weight\_status\_description  order by ows.weight\_status\_description; |
| Group Submission | |
| SQL 1 | select weight\_status,  round(avg(case when gender like '%Male%' then avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_male\_raw,  round(avg(case when gender like '%Female%' then avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_female\_raw  from obesity\_data  group by weight\_status  order by weight\_status; |
| Screenshot |  |
| SQL 2 | select ows.weight\_status\_description as weight\_status,  round(avg(case when og.gender\_name like '%Male%' then op.avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_male\_norm,  round(avg(case when og.gender\_name like '%Female%' then op.avg\_days\_activity\_weekly else null end), 2) as avg\_phy\_act\_days\_female\_norm  from obesity\_patient op  join obesity\_gender og on op.gender\_id = og.gender\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  group by ows.weight\_status\_description  order by ows.weight\_status\_description; |
| Screenshot |  |

# Section 3

|  |  |
| --- | --- |
| Question 11 | **Of the ones that drink frequently, what is the percentage of individuals who are overweight or obese?** |
| Student | ap38689 |
| SQL 1 | select alcohol\_frequency, round(100 \* (count(\*) /  (  select count(\*)  from hc848.obesity\_data  where alcohol\_frequency like 'Frequently'  )), 2) as required\_percent  from hc848.obesity\_data  where (weight\_status like '%Overweight%'  or weight\_status like '%Obesity%')  and (alcohol\_frequency like '%Frequently%')  group by alcohol\_frequency; |
| SQL 2 | select oof.frequency\_description, round(100 \* (count(\*) /  (  select count(\*)  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  where oof.frequency\_description like '%Frequently%'  )), 2) as required\_frequency  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  where (ows.weight\_status\_description like '%Overweight%' or ows.weight\_status\_description like '%Obesity%')  and (oof.frequency\_description like '%Frequently%')  group by oof.frequency\_description; |
| Group Submission | |
| SQL 1 | select alcohol\_frequency, round(100 \* (count(\*) /  (  select count(\*)  from hc848.obesity\_data  where alcohol\_frequency like 'Frequently'  )), 2) as required\_percent  from hc848.obesity\_data  where (weight\_status like '%Overweight%'  or weight\_status like '%Obesity%')  and (alcohol\_frequency like '%Frequently%')  group by alcohol\_frequency; |
| Screenshot |  |
| SQL 2 | select oof.frequency\_description, round(100 \* (count(\*) /  (  select count(\*)  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  where oof.frequency\_description like '%Frequently%'  )), 2) as required\_frequency  from obesity\_patient op  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  where (ows.weight\_status\_description like '%Overweight%' or ows.weight\_status\_description like '%Obesity%')  and (oof.frequency\_description like '%Frequently%')  group by oof.frequency\_description; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 12 | **How likely is (probability) a person to be obese given that they do not eat high caloric food, do not smoke and do not drink alcohol?** |
| Student | pp673 |
| SQL 1 | select round(count(case when weight\_status like '%Obesity%' then 1 end) / count(\*), 2) as req\_probability\_raw  from hc848.obesity\_data  where eat\_high\_caloric\_food like '%no%'  and smoke like '%no%'  and alcohol\_frequency like '%no%'; |
| SQL 2 | select  round((  select count(\*)  from obesity\_patient op  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  join obesity\_boolean obe on op.eat\_high\_caloric\_food\_id = obe.boolean\_id  join obesity\_boolean obs on op.smoke\_id = obs.boolean\_id  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  where ows.weight\_status\_description like '%Obesity%'  and obe.boolean\_value like '%No%'  and obs.boolean\_value like '%No%'  and oof.frequency\_description like '%Never%'  ) / count(\*), 2) as req\_probability\_norm  from obesity\_patient op  join obesity\_boolean obe on op.eat\_high\_caloric\_food\_id = obe.boolean\_id  join obesity\_boolean obs on op.smoke\_id = obs.boolean\_id  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  where obe.boolean\_value like '%No%'  and obs.boolean\_value like '%No%'  and oof.frequency\_description like '%Never%'  group by obs.boolean\_value, obe.boolean\_value, oof.frequency\_description; |
| Group Submission | |
| SQL 1 | select  round((  select count(\*)  from hc848.obesity\_data  where weight\_status like '%Obesity%'  and eat\_high\_caloric\_food like '%no%'  and smoke like '%no%'  and alcohol\_frequency like '%no%'  )/count(\*), 2) as req\_probability\_raw  from hc848.obesity\_data  where eat\_high\_caloric\_food like '%no%'  and smoke like '%no%'  and alcohol\_frequency like '%no%'  group by eat\_high\_caloric\_food, smoke, alcohol\_frequency; |
| Screenshot |  |
| SQL 2 | select round(count(case when ows.weight\_status\_description like '%Obesity%' then 1 end) / count(\*), 2) as req\_probability\_norm  from obesity\_patient op  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  join obesity\_boolean obe on op.eat\_high\_caloric\_food\_id = obe.boolean\_id  join obesity\_boolean obs on op.smoke\_id = obs.boolean\_id  join obesity\_frequency oof on op.alcohol\_frequency\_id = oof.frequency\_id  where obe.boolean\_value like '%No%'  and obs.boolean\_value like '%No%'  and oof.frequency\_description like '%Never%'; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 13 | **Out of the people who monitor their calories daily, what is the percentage of the people that work out?** |
| Student | ar3876 |
| SQL 1 | select ROUND(100\*(count(CASE WHEN AVG\_DAYS\_ACTIVITY\_WEEKLY>0 THEN 1 END)/COUNT(\*)),2) AS MONCALDAILYWORKOUTPERC  from hc848.OBESITY\_DATA WHERE MONITOR\_CALORIES\_DAILY like '%yes%'; |
| SQL 2 | select ROUND(100\*(count(CASE WHEN OP.AVG\_DAYS\_ACTIVITY\_WEEKLY>0 THEN 1 END)/COUNT(\*)),2) AS MONCALDAILYWORKOUTPERC  from OBESITY\_PATIENT OP  JOIN OBESITY\_BOOLEAN OB ON OP.MONITOR\_CALORIES\_DAILY\_ID=OB. BOOLEAN\_ID WHERE MONITOR\_CALORIES\_DAILY\_ID= 'Y'; |
| Group Submission | |
| SQL 1 | select  round(100 \* ((  select count(\*)  from hc848.obesity\_data  where monitor\_calories\_daily like '%yes%'  and avg\_days\_activity\_weekly > 0  ) / count(\*)), 2) as req\_percent\_raw  from hc848.obesity\_data  where monitor\_calories\_daily like '%yes%'  group by monitor\_calories\_daily; |
| Screenshot |  |
| SQL 2 | select  round(100 \* ((  select count(\*)  from obesity\_patient op  join obesity\_boolean ob on op.monitor\_calories\_daily\_id = ob.boolean\_id  where ob.boolean\_value like '%Yes%'  and op.avg\_days\_activity\_weekly > 0  ) / count(\*) ), 2) as req\_percent\_norm  from obesity\_patient op  join obesity\_boolean ob on op.monitor\_calories\_daily\_id = ob.boolean\_id  where ob.boolean\_value like '%Yes%'  group by ob.boolean\_value; |
| Screenshot |  |

|  |  |
| --- | --- |
| Question 14 | **What is the % distribution of obese people based on the numbers of the main meals taken daily and their preference of eating high caloric food?** |
| Student | vv354 |
| SQL 1 | select main\_meals\_daily,  round(100 \* (count(case when eat\_high\_caloric\_food like '%yes%' then 1 end) / (select count(\*) from obesity\_data)), 2) as prefer\_high\_caloric\_food\_pct\_raw,  round(100 \* (count(case when eat\_high\_caloric\_food like '%no%' then 1 end) / (select count(\*) from obesity\_data)), 2) as prefer\_low\_caloric\_food\_pct\_raw  from obesity\_data  where weight\_status like '%Obesity%'  group by main\_meals\_daily  order by main\_meals\_daily; |
| SQL 2 | select op.main\_meals\_daily,  round(100 \* (count(case when ob.boolean\_value like '%Yes%' then 1 end) / (select count(\*) from obesity\_patient)), 2) as prefer\_high\_caloric\_food\_pct\_norm,  round(100 \* (count(case when ob.boolean\_value like '%No%' then 1 end) / (select count(\*) from obesity\_patient)), 2) as prefer\_low\_caloric\_food\_pct\_norm  from obesity\_patient op  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  join obesity\_boolean ob on op.eat\_high\_caloric\_food\_id = ob.boolean\_id  where ows.weight\_status\_description like '%Obesity%'  group by op.main\_meals\_daily  order by op.main\_meals\_daily; |
| Group Submission | |
| SQL 1 | select main\_meals\_daily,  round(100 \* (count(case when eat\_high\_caloric\_food like '%yes%' then 1 end) / (select count(\*) from obesity\_data)), 2) as prefer\_high\_caloric\_food\_pct\_raw,  round(100 \* (count(case when eat\_high\_caloric\_food like '%no%' then 1 end) / (select count(\*) from obesity\_data)), 2) as prefer\_low\_caloric\_food\_pct\_raw  from obesity\_data  where weight\_status like '%Obesity%'  group by main\_meals\_daily  order by main\_meals\_daily; |
| Screenshot |  |
| SQL 2 | select op.main\_meals\_daily,  round(100 \* (count(case when ob.boolean\_value like '%Yes%' then 1 end) / (select count(\*) from obesity\_patient)), 2) as prefer\_high\_caloric\_food\_pct\_norm,  round(100 \* (count(case when ob.boolean\_value like '%No%' then 1 end) / (select count(\*) from obesity\_patient)), 2) as prefer\_low\_caloric\_food\_pct\_norm  from obesity\_patient op  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  join obesity\_boolean ob on op.eat\_high\_caloric\_food\_id = ob.boolean\_id  where ows.weight\_status\_description like '%Obesity%'  group by op.main\_meals\_daily  order by op.main\_meals\_daily; |
| Screenshot | Inserting image... |

|  |  |
| --- | --- |
| Question 15 | **What is the obesity rate at each gender for the individuals who are still considered minor (under 18)?** |
| Student | vv354 |
| SQL 1 | select gender, round(100 \* (count(case when weight\_status like '%Obesity%' then 1 end) / count(\*)) , 2) as obesity\_rate\_raw  from obesity\_data  where age < 18  group by gender; |
| SQL 2 | select og.gender\_name as gender,  round(100 \* (count(case when ows.weight\_status\_description like '%Obesity%' then 1 end) / count(\*)), 2) as obesiry\_rate\_norm  from obesity\_patient op  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  join obesity\_gender og on op.gender\_id = og.gender\_id  where op.age < 18  group by og.gender\_name; |
| Group Submission | |
| SQL 1 |  |
| Screenshot | select gender, round(100 \* (count(case when weight\_status like '%Obesity%' then 1 end) / count(\*)) , 2) as obesity\_rate\_raw  from obesity\_data  where age < 18  group by gender; |
| SQL 2 |  |
| Screenshot | select og.gender\_name as gender,  round(100 \* (count(case when ows.weight\_status\_description like '%Obesity%' then 1 end) / count(\*)), 2) as obesiry\_rate\_norm  from obesity\_patient op  join obesity\_weight\_status ows on op.weight\_status\_id = ows.weight\_status\_id  join obesity\_gender og on op.gender\_id = og.gender\_id  where op.age < 18  group by og.gender\_name; |

# Challenge (extra credit)

|  |  |
| --- | --- |
| Challenge  Question | **What is the overall condition or combination of two conditions with the highest probability of someone being obese (Obesity Type I, II or III)?** |
| Group Submission Only using OBESITY\_DATA table only | |
| SQL | select food\_between\_meals\_frequency,  overweight\_family\_history,  count(\*) as total\_patients,  sum(case when weight\_status like '%Obesity%' then 1 else 0 end) as obese\_patients,  round(sum(case when weight\_status like '%Obesity%' then 1 else 0 end) / count(\*), 2) as obese\_probability,  rank() over(order by round(sum(case when weight\_status like '%Obesity%' then 1 else 0 end) / count(\*), 2) desc) as rank  from obesity\_data  group by overweight\_family\_history,food\_between\_meals\_frequency  order by obese\_probability desc  fetch first 1 rows only; |
| Screenshot |  |
| Comments | We have made a brute force method where we checked each columns and figured out the conditions that are mentioned in the above. We are not sure if this is highest probability. |