```
-- Databricks notebook source

drop table graph;

-- COMMAND ----------

create table graph
using csv
options (path"/FileStore/tables/graph.csv",header "true" );

-- COMMAND ----------

select * from graph limit 10;

-- COMMAND ----------

create table one_degree_target
using csv
options (PATH "/FileStore/tables/one_degree_target.csv", HEADER "TRUE");

-- COMMAND ----------

select * from one_degree_target limit 10;

-- COMMAND ----------

create table people
using csv
options(PATH "/FileStore/tables/people.csv", HEADER "TRUE");

-- COMMAND ----------

select * from people limit 10;

-- COMMAND ----------

create table random_targets
using csv
options(PATH "/FileStore/tables/random_targets.csv", HEADER = "TRUE");

-- COMMAND ----------

select * from random_targets limit 10;
```

-- COMMAND ----------

```
create table recent_purchases
using csv
options(PATH "/FileStore/tables/recent_purchases.csv",HEADER = "TRUE");
```

-- COMMAND ----------

```
select * from recent_purchases limit 10;
```

-- COMMAND ---------- **For first question**----------

```
select count(uid))*100/
      (select count(*)
      from one_degree_target) as clicked_bought
from one_degree_target
where ad_action = "clicked" and buy_action = "bought_car";
```

-- COMMAND ----------

```
select count(uid) from ONE_DEGREE_TARGET;
```

-- COMMAND ----------

```
select count(uid)
from one_degree_target
where ad_action = "clicked" and buy_action = "bought_car";
```

-- COMMAND ----------

```
select count(uid)*100/
      (select count(*)
      from one_degree_target) as clicked_notbought
from one_degree_target
where ad_action = "clicked" and buy_action is null;
```

-- COMMAND ----------

```
select count(uid)
from one_degree_target
where ad_action = "clicked" and buy_action is null;
```

```
-- COMMAND ----------

select count(uid)*100/
      (select count(*)
      from one_degree_target) as clicked_bought
from one_degree_target
where ad_action = "did_not_click" and buy_action = "bought_car";

-- COMMAND ----------

select count(uid)
from one_degree_target
where ad_action = "did_not_click" and buy_action = "bought_car";

-- COMMAND ----------

select count(uid)*100/
      (select count(*)
      from one_degree_target) as notclicke_notbought
from one_degree_target
where ad_action = "did_not_click" and buy_action is null;

-- COMMAND ----------

select count(*)
from one_degree_target
where ad_action = "did_not_click" and buy_action is null;

-- COMMAND ----------

select count(uid)*100/
      (select count(*)
      from random_targets) as notclicked_bought
from random_targets
where ad_action = "did_not_click" and buy_action = "bought_car";

-- COMMAND ----------

select count(uid)
from random_targets
where ad_action = "did_not_click" and buy_action = "bought_car";

-- COMMAND ----------
```

```
select(uid)
from random_targets
where ad_action = "clicked" and buy_action = "bought_car";

-- COMMAND ----------

select count(uid)*100/
  (select count (*) from random_targets)as clicked_notbought
from random_targets
where ad_action = "clicked" and buy_action is null;


-- COMMAND ----------

select count(uid)
from random_tagrets
where ad_action = "clicked" and buy_action is NULL;

-- COMMAND ----------

select count(uid)*100/
  (select count(*) from random_targets) as notclicked_notbought
from random_targets
where ad_action = "did_not_click" and buy_action is null;

-- COMMAND ----------

select count(uid)
from random_targets
where ad_action = "did_not_click" and buy_action is null;

-- COMMAND ----------

select `Family Income Detector` from PEOPLE limit 10;
```

-- COMMAND ----------**For the Third Question**----------

```
with totals as (
  select edges, count(1) as total
  from one_degree_target
  join graph on one_degree_target.uid = graph.sink_uid
  group by 1
)
select p.edges, r.ad_action, r.buy_action, count(1) as value, g.total, count(1) / g.total * 100 as
percentage
from one_degree_target r
join graph p on p.sink_uid = r.uid
join totals g on g.edges = p.edges
where ad_action = "clicked" and buy_action = "bought_car"
group by 1, 2, 3, 5
order by 6
```

-- COMMAND ----------**For the second Question**----------

```
with totals as (
  select gender, count(1) as total
  from one_degree_target
  join people on people.uid = one_degree_target.uid
  group by 1
)
select p.gender, r.ad_action, r.buy_action, count(1) as value, g.total, count(1) / g.total * 100 as
percentage
from one_degree_target r
join people p on p.uid = r.uid
join totals g on g.gender = p.gender
where ad_action = "clicked" and buy_action = "bought_car"
group by 1, 2, 3, 5
order by 6
```

-- COMMAND ----------

```
with totals as (
  select race, count(1) as total
  from one_degree_target
  join people on people.uid = one_degree_target.uid
  group by 1
)
select p.race, r.ad_action, r.buy_action, count(1) as value, g.total, count(1) / g.total * 100 as
percentage
from one_degree_target r
join people p on p.uid = r.uid
join totals g on g.race = p.race
where ad_action = "clicked" and buy_action = "bought_car"
group by 1, 2, 3, 5
order by 6
```

-- COMMAND ----------

```
with totals as (
  select religion, count(1) as total
  from one_degree_target
  join people on people.uid = one_degree_target.uid
  group by 1
)
select p.religion, r.ad_action, r.buy_action, count(1) as value, g.total, count(1) / g.total * 100 as
percentage
from one_degree_target r
join people p on p.uid = r.uid
join totals g on g.religion = p.religion
where ad_action = "clicked" and buy_action = "bought_car"
group by 1, 2, 3, 5
order by 6
```

-- COMMAND ----------

```
with totals as (
  select
  case
  when `Family Income Detector` <= 10000 then "less than 10k"
  when `Family Income Detector` between 10001 and 50000 then "between 10k and 50k"
  when `Family Income Detector` between 50001 and 100000 then "between 50k and 300k"
  when `Family Income Detector` between 100001 and 300000 then "between 100k and 300k"
  when `Family Income Detector` between 300001 and 500000 then "between 300k and 500k"
  when `Family Income Detector` > 500000 then "greater than 500k"
  end as salary, count(1) as total
  from one_degree_target
  join people on people.uid = one_degree_target.uid
  group by 1
),
value as (
  select
  case
  when `Family Income Detector` <= 10000 then "less than 10k"
  when `Family Income Detector` between 10001 and 50000 then "between 10k and 50k"
  when `Family Income Detector` between 50001 and 100000 then "between 50k and 300k"
  when `Family Income Detector` between 100001 and 300000 then "between 100k and 300k"
  when `Family Income Detector` between 300001 and 500000 then "between 300k and 500k"
  when `Family Income Detector` > 500000 then "greater than 500k"
  end as salary, ad_action, buy_action, count(1) as value
  from one_degree_target
  join people on people.uid = one_degree_target.uid
  group by 1, 2, 3
)
select v.salary, v.ad_action, v.buy_action, v.value, t.total, v.value/t.total * 100 as percentage
from value v
join totals t on t.salary = v.salary
where ad_action = "clicked" and buy_action = "bought_car"
group by 1, 2, 3, 4, 5, 6
order by 1, 2, 3, 6
```

```
-- COMMAND ----------
with totals as (
select
case
when date_format(current_date(), "YYYY") - `Birth Year` < 18 then "less than 18"
when date_format(current_date(), "YYYY") - `Birth Year` between 18 and 34 then "between 18
and 34"
when date_format(current_date(), "YYYY") - `Birth Year` between 35 and 50 then "between 35
and 50"
when date_format(current_date(), "YYYY") - `Birth Year` between 51 and 69 then "between 51
and 69"
when date_format(current_date(), "YYYY") - `Birth Year` between 70 and 87 then "between 70
and 87"
when date_format(current_date(), "YYYY") - `Birth Year` >87 then "greater than 87"
end as age,
count(1) as total
from one_degree_target
join people on people.uid = one_degree_target.uid
group by 1
),
value as (
select
case
when date_format(current_date(), "YYYY") - `Birth Year` < 18 then "less than 18"
when date_format(current_date(), "YYYY") - `Birth Year` between 18 and 34 then "between 18
and 34"
when date_format(current_date(), "YYYY") - `Birth Year` between 35 and 50 then "between 35
and 50"
when date_format(current_date(), "YYYY") - `Birth Year` between 51 and 69 then "between 51
and 69"
when date_format(current_date(), "YYYY") - `Birth Year` between 70 and 87 then "between 70
and 87"
when date_format(current_date(), "YYYY") - `Birth Year` >87 then "greater than 87"
end as age, ad_action, buy_action, count(1) as value
from one_degree_target
join people on people.uid = one_degree_target.uid
group by 1, 2, 3
)
select v.age, v.ad_action, v.buy_action, v.value, t.total,
v.value/t.total * 100 as percentage
from value v
join totals t on t.age = v.age
where ad_action = "clicked" and buy_action = "bought_car"
```

```
group by 1,2,3,4,5,6
order by 1,2,3,6

-- COMMAND ----------

with random as (
  select p.*, r.ad_action, r.buy_action
  from random_targets r
  join people p on p.uid = r.uid
),
one_degree as (
  select p.*, o.ad_action, o.buy_action
  from one_degree_target o
  join people p on p.uid = o.uid
)
select
  count(*) as people_in_each_group,
  sum(CASE WHEN r.ad_action = "clicked" THEN 1 ELSE 0 END) as random_clicked,
  sum(CASE WHEN o.ad_action = "clicked" THEN 1 ELSE 0 END) as one_degree_clicked,
  sum(CASE WHEN r.buy_action = "bought_car" THEN 1 ELSE 0 END) as random_bought,
  sum(CASE WHEN o.buy_action = "bought_car" THEN 1 ELSE 0 END) as one_degree_bought
from random r, one_degree o
where r.`Family Income Detector` = o.`Family Income Detector`
and r.gender = o.gender
and r.`Birth Year` = o.`Birth Year`
and r.race = o.race
and r.religion = o.religion
and r.ethnicity = o.ethnicity
```