```
--ODPS_SQL
--*********************************************************************--
--Theme: GG Task
--Description: Analyze the performance of the referral in GG
--Creator: Maya
--Create Date: 2023-03-15 22:54:28
--Modify Date:
--*********************************************************************--
```

1 Please insert the data provided as CSV into tables in an SQL database. Please include SQL queries used throughout the assignment.

1.1 create table in the DB

```
create table referrals (
     id              STRING COMMENT 'id'
     ,created_at     STRING COMMENT 'created_at'
     ,updated_at     STRING COMMENT 'updated_at'
     ,company_id     STRING COMMENT 'company_id'
     ,partner_id     STRING COMMENT 'partner_id'
     ,consultant_id STRING COMMENT 'consultant_id'
     ,status         STRING COMMENT 'status'
     ,is_outbound    STRING COMMENT 'is_outbound'
                          )
comment 'referrals'
;
create table partners (
     id                 STRING COMMENT 'id'
     ,created_at         STRING COMMENT 'created_at'
     ,updated_at          STRING COMMENT 'updated_at'
     ,partner_type        STRING COMMENT 'partner_type'
     ,lead_sales_contact STRING COMMENT 'lead_sales_contact'
                          )
comment 'partners'
;
create table sales_people (
      name      STRING COMMENT 'name'
     ,country STRING COMMENT 'country'
                          )
comment 'sales_people'
;
```

1.2 load data into the tables and check if success

The datasets are not large, so there is no need to check detailed content one by one.

```
select * from referrals;
select count(1) from referrals;
```

```
select * from partners;
select count(1) from partners;


select * from sales_people;
select count(1) from sales_people;
```

2 Use dbt to pre-process the data and output dbt models for analysis. Include appropriate data quality tests and documentation.

I do not have dbt account, so I just use my company's tech stack called DataPhin.

This stage is to transform the raw data to normalized datasets for further analysis, which including the format adjustment, dublicates removement, null value check and unique primary key check.

```
with referrals_new as
(
    select  id
            ,coalesce(from_unixtime(substring(created_at,1,10)),'-') as created_at --change unix nano
timestamp into 'yyyy-mm-dd hh:mm:ss' format
            ,coalesce(from_unixtime(substring(updated_at,1,10)),'-') as updated_at --change unix nano
timestamp into 'yyyy-mm-dd hh:mm:ss' format
            ,coalesce(company_id,'-') as company_id    --null value check
            ,coalesce(partner_id,'-') as partner_id    --null value check
            ,coalesce(consultant_id,'-') as consultant_id --null value check
            ,coalesce(status,'-') as status        --null value check
            ,coalesce(is_outbound,'-') as is_outbound --null value check
    from referrals
    where id is not null --make sure the key colume has value
    group by    id
            ,coalesce(from_unixtime(substring(created_at,1,10)),'-')
            ,coalesce(from_unixtime(substring(updated_at,1,10)),'-')
            ,coalesce(company_id,'-')
            ,coalesce(partner_id,'-')
            ,coalesce(consultant_id,'-')
            ,coalesce(status,'-')
            ,coalesce(is_outbound,'-') --remove duplicates
),
partners_new as
(
 select    id
            ,coalesce(from_unixtime(substring(created_at,1,10)),'-') as created_at --change unix nano
timestamp into 'yyyy-mm-dd hh:mm:ss' format
            ,coalesce(from_unixtime(substring(updated_at,1,10)),'-') as updated_at --change unix nano
timestamp into 'yyyy-mm-dd hh:mm:ss' format
            ,coalesce(partner_type,'-') as partner_type    --null value check
            ,coalesce(lead_sales_contact,'-') as lead_sales_contact     --null value check
```

```
    from partners
    where id is not null --make sure the key colume has value
    group by   id
                ,coalesce(from_unixtime(substring(created_at,1,10)),'-')
                ,coalesce(from_unixtime(substring(updated_at,1,10)),'-')
                ,coalesce(partner_type,'-')
                ,coalesce(lead_sales_contact,'-')    --remove duplicates
),
sales_people_new as
(
 select    coalesce(name,'-') as name       --null value check
           ,coalesce(country,'-') as country   --null value check
   from sales_people
   group by   coalesce(name,'-')
               ,coalesce(country,'-')     --remove duplicates
),
whole_data as
(
    select   a.id as referral_id
             ,a.created_at as referral_created_at
             ,a.updated_at as referral_updated_at
             ,a.company_id
             ,a.partner_id
             ,b.partner_type
             ,b.created_at as partner_created_at
             ,b.updated_at as partner_updated_at
             ,b.lead_sales_contact as partner_sales
             ,c.country as sales_country
             ,a.consultant_id
             ,a.status
             ,a.is_outbound
    from referrals_new a
    left join partners_new b on b.id=partner_id
    left join sales_people_new c on c.name=b.lead_sales_contact
)

select id from referrals_new group by id having count(1)>1; --id is unique
select id from partners_new group by id having count(1)>1; --id is unique
select * from whole_data;
```
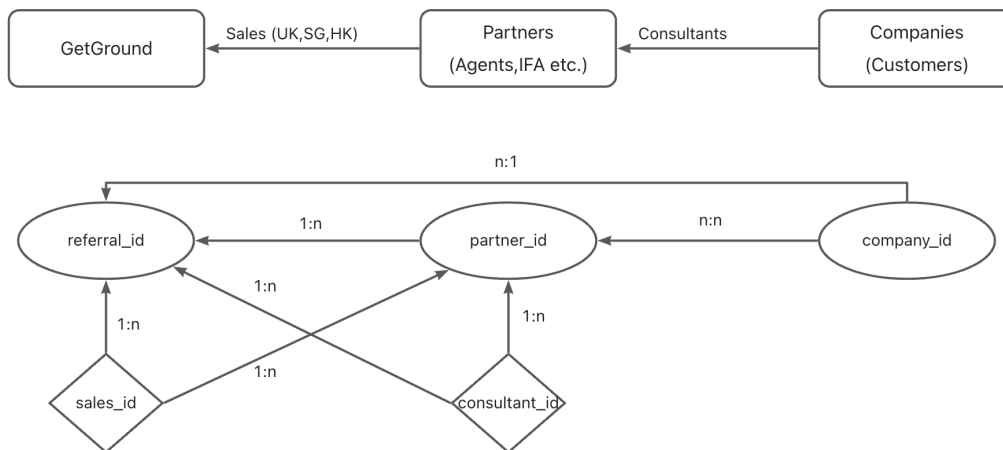
3 Analyze the data using SQL. Be sure to include your investigative thought process, findings, limitations, and assumptions.

Assumptions1: Three datasets are all complete among the provided periods (2020.10-2021.5).)
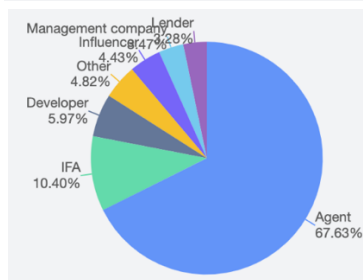
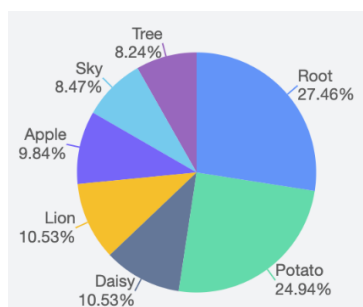## 3.1 Relationships among these subjects

3.2 Partners Information

Findings1: Most partners are agents, which account for 67.63%.

select partner_type,count(id) from partners_new group by partner_type;

Findings2: Root and Potato manage the most customer accounts.

select lead_sales_contact,count(id) from partners_new group by lead_sales_contact;

Findings3: January to April in 2021 got the most partners.

select concat(year(to_date(substring(created_at,1,10),'yyyy-mm-dd')),'.',month(to_date(substring(created_at,1,10),'yyyy-mm-dd'))), count(id) from partners_new group by substring(created_at,1,10);
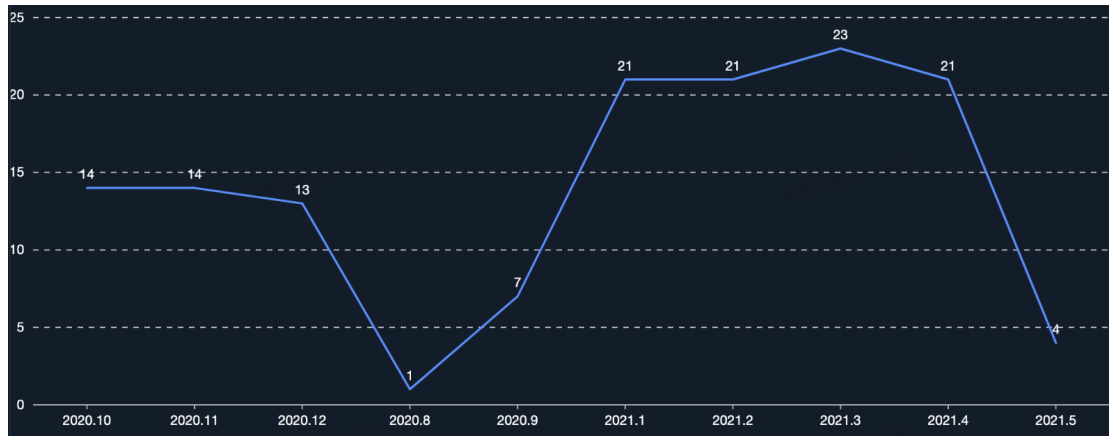
3.3 Sales Information

Findings4: 6 sales are in the UK; 4 sales are in Hong Kong and 2 sales are in Singapore.

select country,count(name) from sales_people_new group by country;

3.4 Referral Information

Findings5: 19.76% of companies is outbound.

select count(distinct company_id),is_outbound from referrals_new group by is_outbound;

Findings6: Partner 7 makes the most referrals for GG and partner 191 and 149 also makes over 100 referrals

select partner_id,count(distinct company_id)from referrals_new group by partner_id;

Findings7: Most referrals spent less than 3 days from pending to successful, while most referrals (>50%) from partner 191/302/360/55/80/30 spent more than 3 days becoming successful.

select partner_id,count(if(datediff<3,id,null))/count(id) as succ_speed_rate
from (
        select DATEDIFF(to_date(substring(updated_at,1,10),'yyyy-mm-
dd'),to_date(substring(created_at,1,10),'yyyy-mm-dd'),'dd') as datediff
        ,partner_id,id
            from referrals_new
            where status='successful'
            group by DATEDIFF(to_date(substring(updated_at,1,10),'yyyy-mm-
dd'),to_date(substring(created_at,1,10),'yyyy-mm-dd'),'dd')
        ,partner_id,id
) t
having count(if(datediff<3,id,null))/count(id)<0.5
group by partner_id;

3.5 Overall Information

Finding8: Most referrals come from agents, which account for 44.22%.
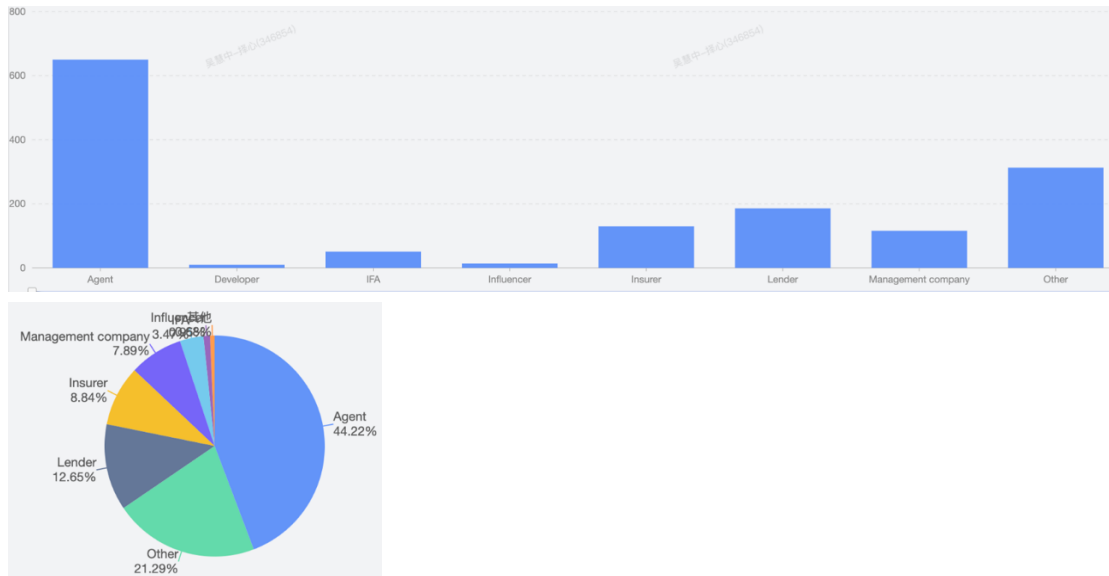
Based on the Finding1, it indicates that other types of partners may have better performance in referrals. And the result shows that partners of insurer perform the best, and each partner brings 130 referrals to GG on average. Management company and lender also brings more referrals.
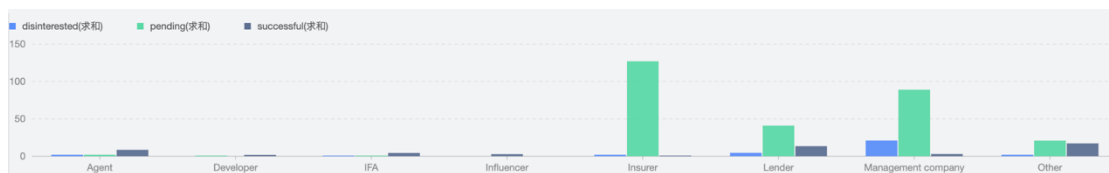
However, most of their referrals are still pending.

	A	B	C
0	partner_type	status	avg_company_cnt
1	IFA	successful	4.363636363636363
2	Agent	disinterested	2.0
3	IFA	disinterested	1.0
4	Agent	pending	2.0
5	Agent	successful	8.56338028169014
6	Other	pending	20.875
7	Other	successful	17.25
8	Management company	successful	3.0
9	IFA	pending	1.0
10	Management company	disinterested	21.0
11	Management company	pending	89.0
	Developer	successful	1.8
13	Insurer	disinterested	2.0
14	Insurer	pending	127.0
15	Insurer	successful	1.0
16	Other	disinterested	2.0
17	Lender	pending	41.0
18	Lender	successful	13.5
19	Lender	disinterested	4.5
20	Influencer	successful	2.8
21	Developer	pending	1.0

```
select partner_type,count(referral_id) from whole_data group by partner_type;
select partner_type,count1/count2
from ( select 'a' as a,partner_type,count(referral_id) as count1 from whole_data group by partner_type )
t1
left join (select 'a' as a,count(referral_id) as count2 from whole_data) t2
on t1.a=t2.a;
select partner_type,status,avg(referral_cnt) as avg_referral_cnt
from (
        select partner_type,partner_id,status,count(distinct referral_id) as referral_cnt
        from whole_data
        group by partner_type,partner_id,status
    ) t
group by partner_type,status;
```

Finding9: 585 referrals on average per person are from accounts whcih are managed by UK sales,
which is the higheset comparing with other countries. And Lion and Tulip perform the best in supporting
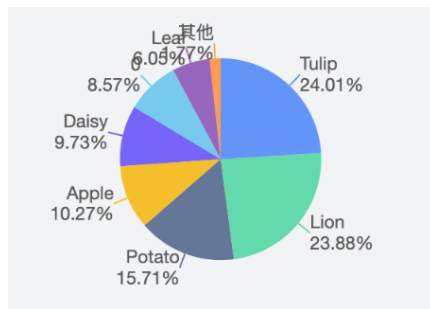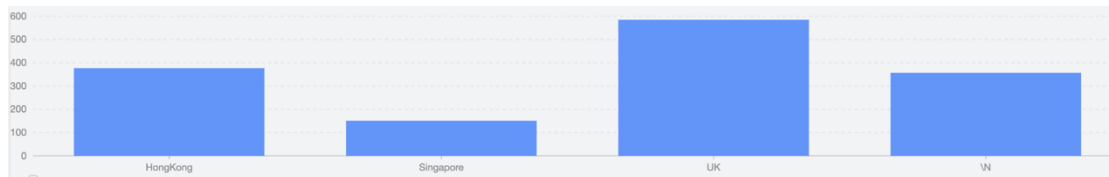their partners in referring customers to GG.

```
select partner_sales,sales_country,avg(referral_cnt) as avg_referral_cnt
from (
        select partner_sales,sales_country,count(distinct referral_id) as referral_cnt
```

```
        from whole_data
        group by partner_sales,sales_country
    ) t
group by partner_sales,sales_country;
```


3.6 Limitation

Limitation1: Not sure if the data is complete during certain periods.

Limitation2: The business background is not clear enough and this is key for the analysis.

Limitation3: How is the commission calculated? It is an important information for analyzing the operational cost.

4 Based on your analysis, how would you recommend GG improve the quality of the analyses we can deliver.

Finding 1&8: Insurer, Management company and Lender have better performance in getting referrals and GG should invest more resources in seeking these types of partners.

Finding 2&9: Although Lion and Tulip do not manage the most partners' accounts, they are the best players in supporting their partners in referring customers to GG. And GG can consider design proper staff motivation scheme to encouraging sales for getting more referrals.

Finding 3: The first season might be the best period in a year for seeking partners.

Finding 4&9: UK may have better market comparing with SG and HK and GG can consider hiring more people in the UK.

Findings5: There still exist great space in outbounding.

Findings6&7: GG can consider designing incentive mechanisms to encourage partners to getting more referrals and making customers sign up with GG soon.