Banking Dataset(Marketing)

19 May 2024 21:51 PM

I have taken this data from the Kaggle for performing some EDA.
I will Write some queries in the SQL as well to analysis the data.

Data Reference:
https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets

Note :- I will EDA in SQL and Python Both, First I will do in Sql I will check all the columns and perform some analysis :

I Will solve the below 10 question on SQL:

- -What is the average balance of customers grouped by their marital status? Additionally, determine how many customers have a zero balance and what the variance is when including and excluding zero balances.
- Which job type has the highest average campaign duration, and what is that duration?
- Find the correlation between having a housing loan and the average balance.
- Determine the percentage of customers in each job type who have taken a loan.
- Identify any significant trends or patterns in the campaign duration over different age groups.
- List all customers who are married and have a tertiary education.
- Count the number of customers who have a default status of 'no' and have a balance greater than 500.
- Find the details of customers who have been contacted on the 5th of May and have a loan.
- Retrieve the number of customers with an unknown job type who have a balance less than 100.
- Select the records of customers who have been contacted more than once (campaign > 1) and their outcome was 'unknown'

What is the average balance of customers grouped by their marital status? Additionally, determine how many customers have a zero balance and what the variance is when including and excluding zero balances.

```
with x as(
            select marital,count(*)as Total_cust,round(avg(cast(balance as float)),2)as Avg_Balance_include_zero from Banking_marketing
            group by marital),
     y as(
            select marital,count(*)as Total_cust_above_zero,round(avg(cast(balance as float)),2)as Avg_Balance_exe_zero from Banking_marketing
            where balance >0
            group by marital)
select x.marital,x.Total_cust,y.Total_cust_above_zero,(x.Total_cust - y.Total_cust_above_zero)as total_cust_zero_bal,
        x.Avg_Balance_include_zero,y.Avg_Balance_exe_zero,
        abs((x.Avg_Balance_include_zero - y.Avg_Balance_exe_zero))as Bal_variance  from x
     join y on x.marital=y.marital
```

Which job type has the highest average campaign duration, and what is that duration?, we see the Age slab as well in this analysis

Where we see the un-employed customer have the maximum campaign duration

```
select top 10 job,avg(cast(duration as float))as Avg_Duration,
            avg(case when Age_bkt_slab ='18 to 30' then cast(duration as float) end) '18 to 30',
            avg(case when Age_bkt_slab ='31 to 45' then cast(duration as float) end) '31 to 45',
            avg(case when Age_bkt_slab ='46 to 60' then cast(duration as float) end) '46 to 60',
            avg(case when Age_bkt_slab ='61 to 80' then cast(duration as float) end) '61 to 80',
            avg(case when Age_bkt_slab ='80 above' then cast(duration as float) end) '80 above'
            from Banking_marketing
group by job
order by Avg_Duration desc
```

Find the correlation between having a housing loan and the average balance.

```
WITH housing_balance AS (
  SELECT
    CASE WHEN housing = 'yes' THEN 1 ELSE 0 END AS housing_loan,
    cast(balance as float)as balance
  FROM Banking_marketing
),
stats AS (
  SELECT
    AVG(housing_loan) AS mean_housing,
    AVG(balance) AS mean_balance,
```

```
    COUNT(*) AS n
  FROM housing_balance
),
deviations AS (
  SELECT
    housing_loan,
    balance,
    (housing_loan - (SELECT mean_housing FROM stats)) AS dev_housing,
    (balance - (SELECT mean_balance FROM stats)) AS dev_balance
  FROM housing_balance
),
squares AS (
  SELECT
    dev_housing,
    dev_balance,
    (dev_housing * dev_housing) AS sq_dev_housing,
    (dev_balance * dev_balance) AS sq_dev_balance,
    (dev_housing * dev_balance) AS prod_dev_housing_balance
  FROM deviations
)
SELECT
  SUM(prod_dev_housing_balance) /
  (SQRT(SUM(sq_dev_housing)) * SQRT(SUM(sq_dev_balance))) AS correlation
FROM squares
```

Determine the percentage of customers in each job type who have taken a loan.

```
WITH main AS (
  SELECT job, COUNT(*) AS total_cnt
  FROM Banking_marketing
  WHERE loan = 'yes'
  GROUP BY job
)
SELECT
  main.*,
  (main.total_cnt*100 /(SELECT SUM(main.total_cnt) FROM main)) AS total_percnt
FROM main
order by total_percnt desc
```

Identify any significant trends or patterns in the campaign duration over different age groups.

```
WITH base_data AS (
  SELECT
    day,
    ROUND(sum(CAST(duration AS float)), 2) AS campaign_duration,
    sum(CASE WHEN Age_bkt_slab = '18 to 30' THEN CAST(duration AS float) END) AS "18 to 30",
    sum(CASE WHEN Age_bkt_slab = '31 to 45' THEN CAST(duration AS float) END) AS "31 to 45",
    sum(CASE WHEN Age_bkt_slab = '46 to 60' THEN CAST(duration AS float) END) AS "46 to 60",
    sum(CASE WHEN Age_bkt_slab = '61 to 80' THEN CAST(duration AS float) END) AS "61 to 80",
    sum(CASE WHEN Age_bkt_slab = '80 above' THEN CAST(duration AS float) END) AS "80 above"
  FROM Banking_marketing
  GROUP BY day
)
SELECT
  day,
  campaign_duration,
  "18 to 30",
  ROUND("18 to 30" / campaign_duration * 100, 2) AS "18 to 30 %",
  "31 to 45",
  ROUND("31 to 45" / campaign_duration * 100, 2) AS "31 to 45 %",
  "46 to 60",
  ROUND("46 to 60" / campaign_duration * 100, 2) AS "46 to 60 %",
  "61 to 80",
  ROUND("61 to 80" / campaign_duration * 100, 2) AS "61 to 80 %",
  "80 above",
  ROUND("80 above" / campaign_duration * 100, 2) AS "80 above %"
FROM base_data
ORDER BY day ASC;
```

List all customers who are married and have a tertiary education.

```
select * from Banking_marketing
```

where marital='married' and education='tertiary'

Count the number of customers who have a default status of 'no' and have a balance greater than 500.

```
SELECT count(*)as total_cnt
FROM Banking_marketing
WHERE "default" = 'no' AND balance > 500;
```

Find the details of customers who have been contacted on the 5th of May and have a loan.

--**Total customer count is = 215 who meet above condition**

```
select count(*) from Banking_marketing
where duration>0 and day='5' and y='yes'
```

Retrieve the number of customers with an unknown job type who have a balance less than 100.

-- T**otal 60 customer have meet the condition**

```
select  * from Banking_marketing
where job='unknown' and balance <100
```

Select the records of customers who have been contacted more than once (campaign > 1) and their outcome was 'unknown'

--**Total customer- 181**

```
select  count(*) from Banking_marketing
where job='unknown' and campaign >1
```

I have solved the above 10 questions. The details are mentioned above for each question.