```sql
create database Project_Banking;
use Project_Banking;

--UNDERSTANDING THE BANK, THE CLIENT BASE & BUSINESS OPERATIONS

--Total number of records in the 'application_data' table
select count(1) as Number_of_records from application_data;

-- The table has more than 3 lakh records of customer credit application data

--The credit types

select name_contract_type,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as percentage
from application_data
group by NAME_CONTRACT_TYPE;

/*
90% of the loans are Cash Loans while around 10% are Revolving Loans.
 There are 2 kinds of credits namely revolving loans and cash loans. Cash loans are credits given upfront with periodical repayments(car loan),
while revolving loans are loans based on usage having a credit limit like Credit Cards. The company seems to pitch more cash loans. Usually these
 structured and secured loans. One can infer that the company is conservative in giving loans since the earning is usually higher in Revolving Loans.
 This however depends on the risk appetite of a bank,competition of other banks,sales strategy, training of employees,the legal regulations, economy and credit worthiness of the customer base.
 */

--Basic gender distribution
select CODE_GENDER,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as percentage
from application_data
group by CODE_GENDER;

/*
65% of the customers are female,34% are males and rest are others.
This bank has a larger female customer base! Few reasons why this could be the case is that
- Demographic conditions in the region- More working females, Higher financial literacy & education,risk taking appetite
- Marketing Strategy - The bank might be targeting more females. One reason could be that the bank has better & loyal female customers.
                        Fraud rate may be lesser in this gender.
- Social Image & Initiatives - The bank could be promoting women empowerment.
- Government Benefits - The bank might be receiving Government Benefits for having a higher female customer base.
- Geographical Conditions - The region where the bank operates might have more
```

```
    females
39
40  */
41
42  --Gender wise credit distribution
43  select name_contract_type,code_gender,count(1) as volume, cast(count
      (name_contract_type)*100.0/sum(count(name_contract_type))over(partition by
      name_contract_type) as decimal(4,2)) as percentage
44  from application_data
45  group by name_contract_type,CODE_GENDER
46
47  /*
48  The Loans are divided at a 2/3 ratio with females on the higher side. This
      means that the strategy of the bank is same across both the product offerings
49  in terms of the gender.
50  */
51
52  --Volume of People who own car & realty wrt. Credit type
53  select name_contract_type
54  ,count(1) as volume
55  ,sum(case when flag_own_car = 'y' then 1 else 0 end) as own_car
56  ,sum(case when flag_own_realty = 'y' then 1 else 0 end) as own_realty
57  from application_data
58  group by name_contract_type
59
60  /* the large volume of cash loans is backed by people having cars and realty,
      which is expected.
61  Also this could be a reason on why the bank is conservative on it's approach.
62  It might be that the client base already has assets which they can use against
      their credit.
63  These credits have lower interest rates and are safer both to the bank and the
      borrower */
64
65  --Income Distribution & Descriptive Statistics wrt. Credit type
66  SELECT distinct name_contract_type AS name_contract_type
67  ,cast(count(1)over(partition by name_contract_type) *100.0/(select count(1)
      from application_data) as decimal(4,2)) as percentage
68  ,cast(avg(amt_income_total)over(partition by name_contract_type) as int) as
      average_income
69  ,min(amt_income_total) over(partition by name_contract_type) as min_income
70  ,max(amt_income_total) over(partition by name_contract_type) as max_income
71  ,PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY amt_income_total) OVER (PARTITION
      BY name_contract_type) AS Median_Income
72  FROM application_data
73
74  /* The average income of clients is equal in both the loan segments.
75  While the Median income is lower in the Revolving Loans type. One reason could
      be that cash loans require security
76  and a higher income level eligibility criteria.
```

```
The Min income in both the loans average around 26000. The Maximum income is much higher incase of cash loans.
With higher credit, banks require higher security.
The Median Income and max income in case of Cash Loans show a huge gap. This gap can be furter analysed by categorizing customers into income_level_flags
*/

--Income & Credit Distribution with Descriptive Statistics wrt. Credit Type
SELECT distinct name_contract_type AS name_contract_type
,cast(count(1)over(partition by name_contract_type) *100.0/(select count(1) from application_data) as decimal(4,2)) as percentage
,cast(avg(amt_income_total)over(partition by name_contract_type) as int) as average_income
,cast(avg(AMT_CREDIT)over(partition by name_contract_type) as int) as average_credit
,min(amt_income_total) over(partition by name_contract_type) as min_income
,min(AMT_CREDIT) over(partition by name_contract_type) as min_credit
,max(amt_income_total) over(partition by name_contract_type) as max_income
,max(AMT_CREDIT) over(partition by name_contract_type) as max_credit
,PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY amt_income_total) OVER (PARTITION BY name_contract_type) AS Median_Income
,PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY AMT_CREDIT) OVER (PARTITION BY name_contract_type) AS Median_Credit
FROM application_data
/*
The Average Credit in Cash Loans is twice the Revolving Loan credits, while the Average & Minimum income is similar.
This supports the bank's conservative approach of dealing credits. The Minimum Credit however is much higher for Revolving Loans.
But the Median Credit is half of Cash Loans.
Also, the bank gives 5 times the income as a revolving loan to the person with lowest income.
This also supports the bank's risk free approach since clients with less assets can avail the loans. The bank pushes for
secured loans.
*/

--Analysis of Goods Amount for which loan is given in case of Cash Loans
SELECT distinct name_contract_type AS name_contract_type
,cast(count(1)over(partition by name_contract_type) *100.0/(select count(1) from application_data) as decimal(4,2)) as percentage
,cast(avg(AMT_GOODS_PRICE)over(partition by name_contract_type) as int) as average_goods_amt
,cast(avg(AMT_CREDIT)over(partition by name_contract_type) as int) as average_credit
,min(AMT_GOODS_PRICE) over(partition by name_contract_type) as min_goods_amt
,min(AMT_CREDIT) over(partition by name_contract_type) as min_credit
,max(AMT_GOODS_PRICE) over(partition by name_contract_type) as max_goods_amt
,max(AMT_CREDIT) over(partition by name_contract_type) as max_credit
```

```sql
,PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY AMT_GOODS_PRICE) OVER (PARTITION BY name_contract_type) AS Median_goods_amt
,PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY AMT_CREDIT) OVER (PARTITION BY name_contract_type) AS Median_Credit
FROM application_data
where NAME_CONTRACT_TYPE = 'Cash Loans'

/*
Usually the credit is higher than the goods amount for which the loan is taken. The reasons why it could be so are
- The Loan might cover additional charges
- The borrower might have a discretion to use the money acc to their needs
- The borrower might be paying off previous dues with a new loan

Overall, the bank does not allow a significant gap between the goods being purchased and the loan amount.
*/

--Basic Income Type distribution
select name_income_type
,count(1) as volume
,cast(count(1)*100.0/sum(count(1))over() as decimal(4,2))as percentage
from application_data
group by NAME_INCOME_TYPE
order by percentage desc

/*
Most of the bank's clients are the Working Class(~52%) & Commercial Associate (23%), followed by Pensioners(18%)
and then State Servants(7%). The bank rarely provides loan to Businessmen, Students and women on Maternity Leave.
This is a vital indicator and confirms that the bank is conservative in nature.

It also signifies that less Education Loans are being Given.
*/

--Basic Housing Type Distribution
select NAME_HOUSING_TYPE
,count(1) as volume
,cast(count(1)*100.0/sum(count(1))over() as decimal(4,2))as percentage
from application_data
group by NAME_HOUSING_TYPE

/*
88% of the clients have their own House/Apartment. This accounts for the high Cash Loans.
*/

--Basic Occupation Distribution
```

```sql
select OCCUPATION_TYPE
,cast(count(1)*100.0/sum(count(1))over() as decimal(4,2))as percentage
from application_data
group by OCCUPATION_TYPE
order by percentage desc

/*
The highest share of 31% of the Occupation type is Null or Unknown. It could happen in cases where the
client has not disclosed their occupation. Incomplete records could be a reason.
On the plus side, there is wide diversity in the bank's client occupation. It caters to both high-level and
lower-income-level clients.
*/

--Region & City Rating Distribution
select REGION_RATING_CLIENT,
count(1) as Frequency,
cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by REGION_RATING_CLIENT

select REGION_RATING_CLIENT_W_CITY,
count(1) as Frequency,
cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by REGION_RATING_CLIENT_W_CITY

/*
The Region & City Ratings are given by the Bank and it seems that 73% of the client are having 2 as the region rating.
Only 10% of the clients are from Rank1 Regions. This could be due to many reasons like
- Competition from other banks
- Less Population in Highly Developed Regions
- Bank's presence might be low in those regions(Distance,Online Reach)
- Pitching to those regions might need more educated/experienced employees(a direct cost to the bank)
*/

-- Education level of the applications

select NAME_EDUCATION_TYPE
,count(1) as volume
,cast(count(1)*100.0/sum(count(1))over() as decimal(4,2))as percentage
from application_data
```

```
194 group by NAME_EDUCATION_TYPE
195 order by percentage desc
196
197 /*
198 Around 72% of the applicants have completed 10th and 12th Standard(in terms of CBSE,ICSE,State Boards,etc). 24% of the applicants have completed Higher Degrees like
199 Graduation,PG,PHD,etc. Around 10000 applicants are dropouts.
200 */
201
202
203 --Family Status of the Bank's Clients
204 select NAME_FAMILY_STATUS
205 ,count(1) as volume
206 ,cast(count(1)*100.0/sum(count(1))over() as decimal(4,2))as percentage
207 from application_data
208 group by NAME_FAMILY_STATUS
209 order by percentage desc
210
211 /*
212 Around 78% of the Bank's clients are/were married.
213 -This could mean that the bank targets a higher age group.
214 -The people who are Single may be sourcing money from other sources for needs - Parents, Relatives, Friends
215 -They may be a customer of another bank
216 -It shows that banking literacy is higher amongst Married people. Additional responsibilities lead to higher need of credit.
217 */
218
219 --Basic Housing Distribution
220 select NAME_HOUSING_TYPE
221 ,count(1) as volume
222 ,cast(count(1)*100.0/sum(count(1))over() as decimal(4,2))as percentage
223 from application_data
224 group by NAME_HOUSING_TYPE
225 order by percentage desc
226
227 /*
228 88% of the clients have their own House/Apartment. This accounts for the high Cash Loans.
229 */
230
231 --Age Brackets of the Clients
232 with age_application as (
233 select
234 case when datediff(year,DATEADd(dd,DAYS_BIRTH,getdate()),GETDATE()) <=25 then '18-25'
235     when datediff(year,DATEADd(dd,DAYS_BIRTH,getdate()),GETDATE()) between 26 and 40 then '26-40'
```

```
236      when datediff(year,DATEADd(dd,DAYS_BIRTH,getdate()),GETDATE()) between 41
           and 55 then '41-55'
237      when datediff(year,DATEADd(dd,DAYS_BIRTH,getdate()),GETDATE()) between 56
           and 65 then '56-65' else '65above' end as age_bracket
238  from application_data)
239  select age_bracket
240  ,count(1) as Frequency
241  ,cast(count(1)*100.0/(select count(1) from application_data)as decimal(4,2)) as
       Percentage
242  from age_application
243  group by age_bracket
244  order by Percentage desc
245
246  /*
247  37% of the clients are between the age 26 amd 55. 20% of the clients are above
       55.
248  Only 4% of the clients are below 25. Like iterated earlier, the need for credit
       comes with more responsibilities.
249  Few people who get really successful early in their career, tend to avail
       credit options to accelerate their growth.
250  Also, very few clients are Students.
251  */
252
253  --Bank's Contact Reach
254  select
255  NAME_CONTRACT_TYPE,
256  cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as
       percentage,
257  sum(FLAG_MOBIL) as mobile_phone,
258  sum(FLAG_EMP_PHONE) as home_phone,
259  sum(FLAG_WORK_PHONE) as work_phone,
260  sum(FLAG_CONT_MOBILE) as phone_reachability,
261  sum(FLAG_EMAIL) as email,
262  cast(sum(FLAG_CONT_MOBILE)*100.0/sum(FLAG_MOBIL) as decimal(4,2)) as
       reachability
263  from application_data
264  group by NAME_CONTRACT_TYPE
265
266  /*
267  The contact reach is around 100% for Cash Loans and 99% for Revolving Loans
268  This is great sign about the bank's scrutiny and loan processing operations.
       The numbers are probably verified
269  and hence the reachability is very high. This gives less scope for fraud and
       ensures timely payment.
270  */
271
272  --Contacts Availability
273  with contact_data as
274  (select
```

```
275  case when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =3 then 'All Contacts
       Available'
276  when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =2 then 'Two Contacts Available'
277  when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =1 then '1 Contact Available'
278  else 'No Contact Available' end as contacts_provided
279  from application_data)
280  select contacts_provided,
281  count(1) as Frequency,
282  cast(count(1)*100.0/(select count(1) from contact_data) as decimal(4,2)) as
       percentage
283  from contact_data
284  group by contacts_provided
285
286  /*
287  Around 62% of the Clients have provided 2 Contacts, and 19% have given either 1
       or all contacts.
288  There is no client without any contact. The documentation seems clearly
       executed.
289  */
290
291  --Documents Submission Analysis
292  with Documents_data as
293  (select
294  case when FLAG_DOCUMENT_2+FLAG_DOCUMENT_3+FLAG_DOCUMENT_4+FLAG_DOCUMENT_5
       +FLAG_DOCUMENT_6+FLAG_DOCUMENT_7+FLAG_DOCUMENT_8+FLAG_DOCUMENT_9
       +FLAG_DOCUMENT_10+FLAG_DOCUMENT_11+FLAG_DOCUMENT_12+FLAG_DOCUMENT_13
       +FLAG_DOCUMENT_14+FLAG_DOCUMENT_15+FLAG_DOCUMENT_16+FLAG_DOCUMENT_17
       +FLAG_DOCUMENT_18+FLAG_DOCUMENT_19+FLAG_DOCUMENT_20+FLAG_DOCUMENT_21
295  between 15 and 20 then '15-20 Documents Available'
296  when FLAG_DOCUMENT_2+FLAG_DOCUMENT_3+FLAG_DOCUMENT_4+FLAG_DOCUMENT_5
       +FLAG_DOCUMENT_6+FLAG_DOCUMENT_7+FLAG_DOCUMENT_8+FLAG_DOCUMENT_9
       +FLAG_DOCUMENT_10+FLAG_DOCUMENT_11+FLAG_DOCUMENT_12+FLAG_DOCUMENT_13
       +FLAG_DOCUMENT_14+FLAG_DOCUMENT_15+FLAG_DOCUMENT_16+FLAG_DOCUMENT_17
       +FLAG_DOCUMENT_18+FLAG_DOCUMENT_19+FLAG_DOCUMENT_20+FLAG_DOCUMENT_21
297  between 10 and 14 then '10-14 Documents Available'
298  when FLAG_DOCUMENT_2+FLAG_DOCUMENT_3+FLAG_DOCUMENT_4+FLAG_DOCUMENT_5
       +FLAG_DOCUMENT_6+FLAG_DOCUMENT_7+FLAG_DOCUMENT_8+FLAG_DOCUMENT_9
       +FLAG_DOCUMENT_10+FLAG_DOCUMENT_11+FLAG_DOCUMENT_12+FLAG_DOCUMENT_13
       +FLAG_DOCUMENT_14+FLAG_DOCUMENT_15+FLAG_DOCUMENT_16+FLAG_DOCUMENT_17
       +FLAG_DOCUMENT_18+FLAG_DOCUMENT_19+FLAG_DOCUMENT_20+FLAG_DOCUMENT_21
299  between 5 and 9 then ' 5-9 Documents Available'
300  else 'Less than 5 Documents Available' end as Documents_provided
301  from application_data)
302  select Documents_provided,
303  count(1) as Frequency,
304  cast(count(1)*100.0/(select count(1) from documents_data) as decimal(5,2)) as
       percentage
305  from documents_data
306  group by Documents_provided
```

```
307
308 /*
309 In Terms of Documents, Upto 4 Documents were procured at max(100%). These
      documents vary from loan to loan.
310 This could be a good sign in the sense that the bank takes less documentation
      before providing credit.
311 A point to check would be that all the necessary information is collected.
      While less paperwork and online documentation
312 is a plus point, the bank should ensure that no information is missed.
      Occupation details are clearly not part of this
313 check(Again it depends on the loan type). Would be a plus if most of it is
      digitised.
314 */
315
316 --Basic Loan Application Day Analysis
317 select WEEKDAY_APPR_PROCESS_START,
318 cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as
      Percentage
319 from application_data
320 group by WEEKDAY_APPR_PROCESS_START
321 order by Percentage desc
322
323 /*
324 More clients prefer applying for credit on weekdays(17% across all weekdays).
325 Few Clients(11%) applied for credit on Saturdays. The banks are usually closed
      on each alternative Saturday.
326 It could also indicate that the clients are not using online channels(Need to
      analyze the sales channel).
327 It could also indicate that the clients are busy with their household chores,
      family time, leisure,etc.
328 */
329
330
331
332
333 -- PART 3 - TARGET VARIABLE & RISK ANALYSIS
334
335 --Overall Analysis of Credit enquiries on the Clients
336
337 /*
338 In general the banks check the credit profile of a client as a whole. There are
      multiple factors which affect the Cibil Score of an individual.
339 Credit Enquiry is just one of them. These Enquiries are of two types.
340 Examples - Soft Enquiry - Employer checking your credit report, Hard Enquiry -
      Bank checking your credit report for approving credits
341 It is assumed that these are Hard Enquiries.
342 */
343 --- 1 Year before the application
344
```

```sql
select AMT_REQ_CREDIT_BUREAU_YEAR
,count(1) as Frequency
,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by AMT_REQ_CREDIT_BUREAU_YEAR
order by percentage desc

/*
43% of Loan Applications come from clients having 0 or 1 cibil checks & 16% from clients having 2 cibil checks. This is a decent sign that the client does not seem to be risky. This could be further analyzed by looking at their cibil reports for 2 years.
20% of clients have more than 2 enquiries in 1 year. This is further analyzed below by looking at their quarterly and monthly enquiries.
13.5% values are null which i assume are the clients having no credit history/ taking credit for the 1st time. This depends on multiple factors like the bank's strategy, legal implications, client relationship(might be a customer having deposits),etc.
Past behavior of clients in that geographical locations need to be checked in order to know if this is risky sign or not. Macro changes in economy(drop in interest rates,Increase in taxes,etc) could also affect this factor.
*/

--- 1 Quarter before the application

select top 5 AMT_REQ_CREDIT_BUREAU_QRT
,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by AMT_REQ_CREDIT_BUREAU_QRT
order by percentage desc

/*
While there were around 23% clients having 1 enquiry in 1 year, 20% clients having 0 enquries in 1 year, and 20% clients having more than 2 enquries in 1 year,
70% of clients among the applicants had 0 enquries in the last 3 months. The same 13.5% clients have no history, 11% clients have 1 and 4.5% clients have 2 quarterly enquiries.
*/

--- 1 Month before the application

select top 5 AMT_REQ_CREDIT_BUREAU_MON
,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by AMT_REQ_CREDIT_BUREAU_MON
```

```
order by percentage desc

/*
Out of the above mentioned enquiry situations, the monthly enquiry is on the safer side as well. 72% clients have 0 enquiry, 13.5% have no credit history,
10% have 1 enquiry and 1.75% have 2 enquiries.
*/

--- 1 Week before the application

select top 5 AMT_REQ_CREDIT_BUREAU_WEEK
,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by AMT_REQ_CREDIT_BUREAU_WEEK
order by percentage desc

/*
83% clients have no enquiries within a week of their application. 2.67% have 1 enquiry.
*/

--- 1 Day before the application

select top 5 AMT_REQ_CREDIT_BUREAU_DAY
,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by AMT_REQ_CREDIT_BUREAU_DAY
order by percentage desc

/*
Around 86% of the clients have no enquiries made on the same day.13.5% clients have no history.
*/

--- 1 Hour before the application

select top 5 AMT_REQ_CREDIT_BUREAU_HOUR
,cast(count(1)*100.0/(select count(1) from application_data) as decimal(4,2)) as Percentage
from application_data
group by AMT_REQ_CREDIT_BUREAU_HOUR
order by percentage desc

/*
Around 86% of the clients have no enquiries made on the same hour.13.5% clients have no history.
*/
```

```
420
421 /*
422 The bank takes a decision based on multiple factors. Here is what i can infer from this particular one.
423 - The bank takes a safe call, being on the conservative side, while choosing whose application to accept. On surface, it appears like it favors applications
424 from users having no credit enquiries within a month. This could mean that the users are loyal to this bank on their services. Few customers who have enquiries within a week
425 or day or hour could be facing some crisis or emergency. They could also be lacking the knowledge of the impact of these credit checks.
426 - An Analysis on the individual applications will provide more clarity about this.
427 */
428
429 --Basic enquiry averages
430 select avg(AMT_REQ_CREDIT_BUREAU_HOUR) as avg_hour_enquiry
431 ,avg(AMT_REQ_CREDIT_BUREAU_DAY ) as avg_day_enquiry
432 ,avg(AMT_REQ_CREDIT_BUREAU_WEEK) as avg_week_enquiry
433 ,avg(AMT_REQ_CREDIT_BUREAU_MON ) as avg_month_enquiry
434 ,avg(AMT_REQ_CREDIT_BUREAU_QRT ) as avg_quarter_enquiry
435 ,avg(AMT_REQ_CREDIT_BUREAU_YEAR) as avg_year_enquiry
436 from application_data
437
438 --Analysis of individual applications based on the credit enquiries
439
440
441 with enquiry_table as
442 (select
443 case when AMT_REQ_CREDIT_BUREAU_YEAR is null then 'No Credit History'
444 when AMT_REQ_CREDIT_BUREAU_YEAR = 0 then 'No Enquiry in the past year'
445 when AMT_REQ_CREDIT_BUREAU_QRT = 0 then 'Had Enquiries within the year'
446 when AMT_REQ_CREDIT_BUREAU_MON = 0 then 'Had Enquiries within the quarter'
447 when AMT_REQ_CREDIT_BUREAU_WEEK = 0 then 'Had Enquiries within the month'
448 when AMT_REQ_CREDIT_BUREAU_DAY = 0 then 'Had Enquiries within the week'
449 when AMT_REQ_CREDIT_BUREAU_HOUR = 0 then 'Had Enquiries within the day' end as Enquiry_Status
450 from application_data)
451 select Enquiry_Status
452 ,count(Enquiry_Status) as Frequency
453 ,cast(count(Enquiry_Status)*100.0/(select count(1) from enquiry_table)as decimal(4,2)) as Percentage
454 from enquiry_table
455 group by Enquiry_Status
456 order by Percentage desc
457 /*
458 Most clients had enquiries only within a year or no enquiry at all(around 86%). This shows that there is less risk for the bank. Their credit history could
```

```
    tell more about this assumption though.
*/

with default_scope as
(select isnull(cast(DEF_60_CNT_SOCIAL_CIRCLE*100.0/NULLIF
  (OBS_60_CNT_SOCIAL_CIRCLE,0) as decimal(5,2)),0) as Percentage
from application_data)
,risk_scope as
(select
case when Percentage=100 then 'Very High Risk'
when Percentage between 75 and 99 then 'High Risk'
when Percentage between 50 and 74 then 'Moderate Risk'
when Percentage between 25 and 49 then 'Low Risk'
when Percentage <25 then 'Very Low Risk' end as Risk_category_60_Days
from default_scope)
select Risk_category_60_Days,
count(1) as Frequency,
cast(count(1)*100.0/(select count(1) from risk_scope) as decimal(5,2)) as
  Percentage
from risk_scope
group by Risk_category_60_Days
order by Percentage desc

with default_scope as
(select isnull(cast(DEF_30_CNT_SOCIAL_CIRCLE*100.0/NULLIF
  (OBS_30_CNT_SOCIAL_CIRCLE,0) as decimal(5,2)),0) as Percentage
from application_data)
,risk_scope as
(select
case when Percentage=100 then 'Very High Risk'
when Percentage between 75 and 99 then 'High Risk'
when Percentage between 50 and 74 then 'Moderate Risk'
when Percentage between 25 and 49 then 'Low Risk'
when Percentage <25 then 'Very Low Risk' end as Risk_category_30_Days
from default_scope)
select Risk_category_30_Days,
count(1) as Frequency,
cast(count(1)*100.0/(select count(1) from risk_scope) as decimal(5,2)) as
  Percentage
from risk_scope
group by Risk_category_30_Days
order by Percentage desc


/*
92% of the applications look to be of low risk based on the social surroundings
   default history in the last 60 days.
This means that the geographical region is good to do business. The people from
   that region have made timely payments, defaults not exceeding 60dpd.
```

```
around 3% clients tend to be highly risky. Around 9421, customers to be precise. 6760 clients have moderate risk.
Overall, the individual behaviors need to be given more weightage while approving applications even though banks do have specific insights about regions.
*/


with default_scope as
(select target, isnull(cast(DEF_30_CNT_SOCIAL_CIRCLE*100.0/NULLIF(OBS_30_CNT_SOCIAL_CIRCLE,0) as decimal(5,2)),0) as Percentage
from application_data)
,risk_scope as
(select target,
case when Percentage=100 then 'Very High Risk'
when Percentage between 75 and 99 then 'High Risk'
when Percentage between 50 and 74 then 'Moderate Risk'
when Percentage between 25 and 49 then 'Low Risk'
when Percentage <25 then 'Very Low Risk' end as Risk_category_30_Days
from default_scope)
select case when target = 0 then 'Never had Payment Difficulties'
else 'Had Payment Difficulties' end as Target
,Risk_category_30_Days
,count(1) as Frequency
,cast(count(1)*100.0/(select count(1) from risk_scope) as decimal(5,2)) as Percentage
from risk_scope
group by case when target = 0 then 'Never had Payment Difficulties'
else 'Had Payment Difficulties' end, Risk_category_30_Days
order by Target


/*
Around 7% customers who are Very Low Risk based on the social surrounding's 30 days payment default history have had Payment Difficulties.
This is the most important bracket according to me. These are the clients who need to be studied more. A deeper dive on the client demographics is crucial to understand this.
Proper meetings with the Debt Managers and other heads of the Collection team will reveal the reason on why the clients defaulted. Maybe they had an emergency, maybe the collection
method was not appropriate. It could also happen that they changed their address or they could not be contacted via email or cell.
For the clients who never had any Payment Difficulties, proper customer service, cross-product selling, long-term relationship building and proper customer service is the key.
*/

--Deeper analysis on the Contact reach for clients who had payment difficulties
```

```
      but were from the Very Low Risk social surroundings
537
538 with default_scope as
539 (select target
540 ,case when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =3 then 'All Contacts
      Available'
541 when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =2 then 'Two Contacts Available'
542 when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =1 then '1 Contact Available'
543 else 'No Contact Available' end as contacts_provided
544 , isnull(cast(DEF_30_CNT_SOCIAL_CIRCLE*100.0/NULLIF(OBS_30_CNT_SOCIAL_CIRCLE,0)
      as decimal(5,2)),0) as Percentage
545 from application_data)
546 ,risk_scope as
547 (select target,contacts_provided,
548 case when Percentage=100 then 'Very High Risk'
549 when Percentage between 75 and 99 then 'High Risk'
550 when Percentage between 50 and 74 then 'Moderate Risk'
551 when Percentage between 25 and 49 then 'Low Risk'
552 when Percentage <25 then 'Very Low Risk' end as Risk_category_30_Days
553 from default_scope)
554 ,risk_based_on_contact_reach as
555 (select case when target = 0 then 'Never had Payment Difficulties'
556 else 'Had Payment Difficulties' end as Target
557 ,contacts_provided
558 ,Risk_category_30_Days
559 ,count(1) as Frequency
560 ,cast(count(1)*100.0/(select count(1) from risk_scope) as decimal(5,2)) as
      Percentage
561 from risk_scope
562 group by case when target = 0 then 'Never had Payment Difficulties'
563 else 'Had Payment Difficulties' end, Risk_category_30_Days,contacts_provided)
564 select Target,contacts_provided,
565 Risk_category_30_Days,
566 Frequency,
567 cast(Frequency*100.0/sum(frequency)over() as decimal(5,2)) as Percentage
568 from risk_based_on_contact_reach
569 where Target = 'Had Payment Difficulties' and Risk_category_30_Days = 'Very Low
      Risk'
570 order by Percentage desc
571
572 /*
573 Out of the clients who have had payment difficulties and were from Very Low
      Risk regions, All contacts were available for around 24% clients.
574 64% clients have provided 2 contacts and 12% clients have provided only 1
      contact. The team needs to get access of more contact details for these two
      classes of clients.
575 There could be family relatives of these clients whom the bank can contact.
      Ofcourse it is done only in extreme cases. Usually it is done for clients
      having more than 90-120dpd or Bucket3-4.
```

```
Further Analysis needs to be done whether the client lives in the given city or not. Also, an assessment of the credit collection team needs to done.
All changes made in the collection strategy should be analyzed. Redundant changes should be overruled.
*/

--Integrating the detail of whether the address in the document matches where the client actually lives and still had payment difficulties

with default_scope as
(select target
,case when REG_REGION_NOT_LIVE_REGION = 1 then 'Address Mismatch' else 'Address Match' end as Address_city_match
,case when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =3 then 'All Contacts Available'
when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =2 then 'Two Contacts Available'
when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =1 then '1 Contact Available'
else 'No Contact Available' end as contacts_provided
, isnull(cast(DEF_30_CNT_SOCIAL_CIRCLE*100.0/NULLIF(OBS_30_CNT_SOCIAL_CIRCLE,0) as decimal(5,2)),0) as Percentage
from application_data)
,risk_scope as
(select target,contacts_provided,Address_city_match,
case when Percentage=100 then 'Very High Risk'
when Percentage between 75 and 99 then 'High Risk'
when Percentage between 50 and 74 then 'Moderate Risk'
when Percentage between 25 and 49 then 'Low Risk'
when Percentage <25 then 'Very Low Risk' end as Risk_category_30_Days
from default_scope)
,risk_based_on_contact_reach as
(select case when target = 0 then 'Never had Payment Difficulties'
else 'Had Payment Difficulties' end as Target
,Address_city_match
,contacts_provided
,Risk_category_30_Days
,count(1) as Frequency
,cast(count(1)*100.0/(select count(1) from risk_scope) as decimal(5,2)) as Percentage
from risk_scope
group by case when target = 0 then 'Never had Payment Difficulties'
else 'Had Payment Difficulties' end,
Risk_category_30_Days,contacts_provided,Address_city_match)
select Target,contacts_provided,
Address_city_match,
Risk_category_30_Days,
Frequency,
cast(Frequency*100.0/sum(frequency)over() as decimal(5,2)) as Percentage
from risk_based_on_contact_reach
where Target = 'Had Payment Difficulties' and Risk_category_30_Days = 'Very Low
```

```
      Risk'
order by Percentage desc

/*
Around 2% Cases had an address mismatch, while having the contact details.
  Although it is a tiny fraction of the whole, it should still be assessed by
  the debt managers.
The underlying reasons for their payment difficulties could be unavailability
  of funds, lack of contingency fund, or a typical pay in the beginning and
  then default kind of scenario.
*/

--Integrating the previous application data

with data as
(select NAME_CONTRACT_STATUS
from application_data a
join previous_application p on a.SK_ID_CURR = p.SK_ID_CURR)
select NAME_CONTRACT_STATUS,
count(1) as frequency,
cast(count(1)*100.0/(select count(1) from data) as decimal(5,2)) as Percentage
from data
group by NAME_CONTRACT_STATUS

with data as
(select p.NAME_CONTRACT_TYPE
from application_data a
join previous_application p on a.SK_ID_CURR = p.SK_ID_CURR)
select NAME_CONTRACT_TYPE,
count(1) as frequency,
cast(count(1)*100.0/(select count(1) from data) as decimal(5,2)) as Percentage
from data
group by NAME_CONTRACT_TYPE

with credit_data as
(select
case when AMT_APPLICATION between 0 and 500000 then 'Very Low Amount'
when AMT_APPLICATION between 500001 and 1000000 then 'Low Amount'
when AMT_APPLICATION between 1000001 and 1500000 then 'Moderate Amount'
when AMT_APPLICATION between 1500001 and 2000000 then 'High Amount'
else 'Very High Amount' end as prev_credits
from application_data a
join previous_application p on a.SK_ID_CURR = p.SK_ID_CURR)
select prev_credits,
count(1) as frequency,
cast(count(1)*100.0/(select count(1) from credit_data) as decimal(5,2)) as
  Percentage
from credit_data
group by prev_credits
```

```sql

select SK_ID_CURR
,count(sk_id_prev) as previous_applications
,cast(sum(case when NAME_CONTRACT_STATUS = 'approved' then 1 else 0 end)*100.0/count(SK_ID_PREV) as decimal(5,2))  as application_approval_rate
from previous_application
group by SK_ID_CURR
having cast(sum(case when NAME_CONTRACT_STATUS = 'approved' then 1 else 0 end)*100.0/count(SK_ID_PREV) as decimal(5,2)) =100.0


with prev_app_data as
(select SK_ID_CURR
,count(sk_id_prev) as previous_applications
,cast(sum(case when NAME_CONTRACT_STATUS = 'approved' then 1 else 0 end)*100.0/count(SK_ID_PREV) as decimal(5,2))  as application_approval_rate
from previous_application
group by SK_ID_CURR
having cast(sum(case when NAME_CONTRACT_STATUS = 'approved' then 1 else 0 end)*100.0/count(SK_ID_PREV) as decimal(5,2)) =100.0)
select top 15 p.*,a.NAME_INCOME_TYPE,a.NAME_EDUCATION_TYPE,OCCUPATION_TYPE
,case when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =3 then 'All Contacts Available'
when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =2 then 'Two Contacts Available'
when FLAG_MOBIL+FLAG_EMP_PHONE+FLAG_WORK_PHONE =1 then '1 Contact Available'
else 'No Contact Available' end as contacts_provided
from prev_app_data p
join application_data a on p.SK_ID_CURR = a.SK_ID_CURR
order by previous_applications desc



select distinct sk_id_curr
from previous_application
where SK_ID_CURR not in (select SK_ID_CURR from application_data)

select distinct NAME_CONTRACT_STATUS
from previous_application


/*
INSIGHTS & RECOMMENDATIONS
- The bank should try to source more Revolving Loans. This can be done by... This could impact in..
- Provide more loans to Businessmen
-Targeting more Single person could give banks more income. These are the customers with whom banks can build a long term relationship
and provide products at every stage of life. Ofcourse this comes with a higher
```

```
      risk, but an evaluation of current Single clients could
702  reveal the credit behavior of this class.
703  - Reach out to the addresses of the clients whose contact info is unreachable.
704  - Occupation details are missing for more than 31.35% of the clients.The bank
       should reach out and collect more information about it.
705  This not only ensures more security, it also gives the bank a chance to pitch
       more products according to the client's occupation.
706  - Reach out to more Occupations like HR Staff,IT Staff and Realty Agents.
707  - Train employees/agents to reach out to Tier 1 Regions. Need to penetrate and
       investigate the reasons on why the reach is so low on Tier 1 & 3 Regions.
708   One of the most effective way is to have periodical meetings with the
        Executives managing the Sales Channels. They work on ground level and can
        say the correct
709   reason. Also, doing so empowers & motivates them that the upper management
        takes their ideas & it makes them feel important and needed.
710  - Reach out to Students or the young age group by tieing up with Universities,
       Colleges & other Online/Offline Education Institutes.
711  - Maintain the current volume of Sales Programs/Strategies on
       regions,occupations,classes where there is high application rate.
712  - More analysis is required on the 4 stages  - Pre-Transaction,During
       Transaction, Post Transaction & Renewal.
713  - Target Low Risk Customers as well. Tailor made solutions for these buckets
       could prove fruitful for the business. Cross product targeting to Low & Very
       Low Risk
714    classes, tie ups with their organisations(if any) and building long term
         relationships is the key for a stable & profitable business.
715    Deeper Analysis on High Risk & Moderate Risk Clients needs to be done.
         Ofcourse, the quantum of profit from these customers needs to be taken into
          consideration.
716    A Very Low Risk client giving less revenue might be less preferable than a
         Moderate Risky client giving more revenue.
717  - A lot of the bank's revenue depends on how the Credit Collection team
       functions. Proper methodology and action on the ground level ensures timely
       payment collection.
718    Periodical training of debt managers,collection agents,third party vendors
         needs to be done to deal with cases where the contact details are available
          and the social
719    surroundings have Very Low Risk in terms of Payment, but the client has
         defaulted.
720    Also, harsh customer service or debt collection methods can hurt the brand
         image in the mind of the client and in the surrounding(long -term). Proper
         check needs to be taken
721    to ensure that the methods are strict but not overly harsh.
722  - The bank needs to provide the clients with the proper information about the
       effects a default can have on the credit score and the future difficulties
       the client could be facing.
723    There could be instances where the debt managers are too rigid with the
         collection while they should be educating the customers about the
         consequences of such behavior.
```

```
- The bank could enquire about the persons who were accompanying the client during the application. The employees at the bank should be well trained to build knowledge about that
  person. This increases reliability on the client who is applying for credit as well as gives an opportunity to pitch products to the companion.
- There is a need to sit down with the people working on ground level and providing them with the info of the analysis. Integrating these minute details could be really fruitful for any
  organisation. Banking as a sector is highly personalised. It becomes unavoidable to take in account these intricate details and apply them in the day to day operations.
  Ex - Finding that a person has incomplete education, could mean that they started a venture. Although the bank could have details about the person's org, a 5 min conversation of the relationship
  manager with the client about his journey from being a dropout to starting his own venture could have a really positive outlook.
- Although it is cheaper for a bank to maintain current customers than acquiring new ones, it should try to target more clients who have completed their higher education. A large chunk
  of clients have only completed secondary education.
*/

/*
CHALLENGES ON RESEARCH
-The Organization type description was not clear. Terms like 'Business Entity Type 1', 'Industry Type 1' was vague
-Application Date is absent, no analysis could be performed in that aspect. We could not ascertain the increase and decrease in applications over various periods.
It is important to know the peak seasons. Usually the need for credit arrives when there is shortage of money. Month wise, it is the 3rd week of a month. This is the time when the need of credit arrives
due to unplanned expenditures or increased spending. Year wise, people tend to need more credit during the 3rd & 4th Quarter. This is the peak time for retail shopping.
-Rural & Urban Segments could not be analyzed since it was not clear from the data.
-The same customer might have multiple applications
-The enquiries made on the client's credit report to the credit buraeu do not highlight which banks enquired about the client. An analysis of that data could reveal
whether it was this bank or multiple banks involved.
-The Quantum of Revenue is missing in these applications. It is a crucial aspect of analysis.
*/

/*
CHALLENGES ON THE BANK
- In general there is a fall in NPA in India, which is a good sign. It now
```

```
      remains as a challenge on the bank's end to take advantage of this factor
      while facing competition from other players.
- There is minimal control over Interest Rates. It is a question of marketing.
- To increase the profitability,
On the revenue side, the bank needs to either increase it's number of clients
      or it's revenue charges(annual fees,transaction fees,etc).
On the cost side, the bank needs to decrease it's fixed/variable costs. Fixed
      costs like Rent, Maintenance, Employee Salary,etc need to be checked.
      Variable costs include interests on deposits, customer handing costs, etc.
*/

```