

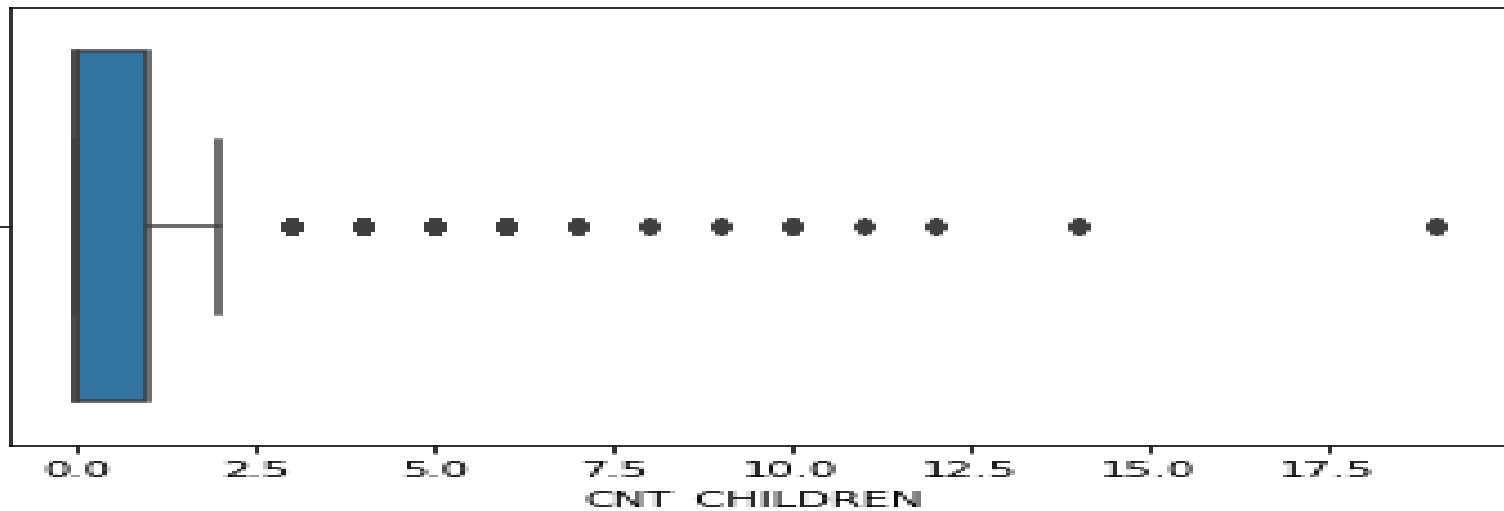
# EDA CASE STUDY

-By Sumit Chakraborty ([sumitch4u@gmail.com](mailto:sumitch4u@gmail.com))

-Akarsh Tyagi ([13akarsh@gmail.com](mailto:13akarsh@gmail.com))

# OUTLIER ANALYSIS ON (CNT\_CHILDREN)

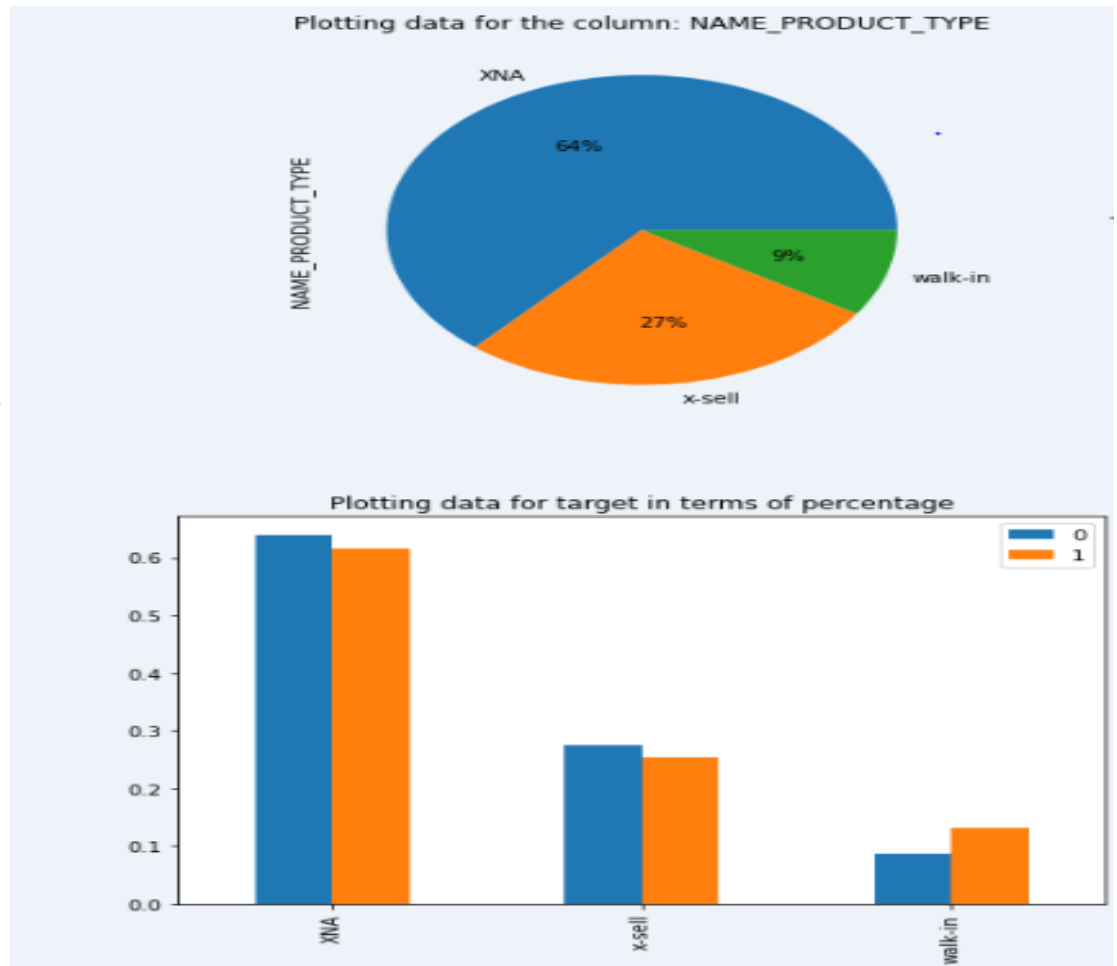
The CNT\_CHILDREN variable has much outliers as many customers are having more than 5 children which is not possible in current circumstances.



# UNIVARIATE ANALYSIS

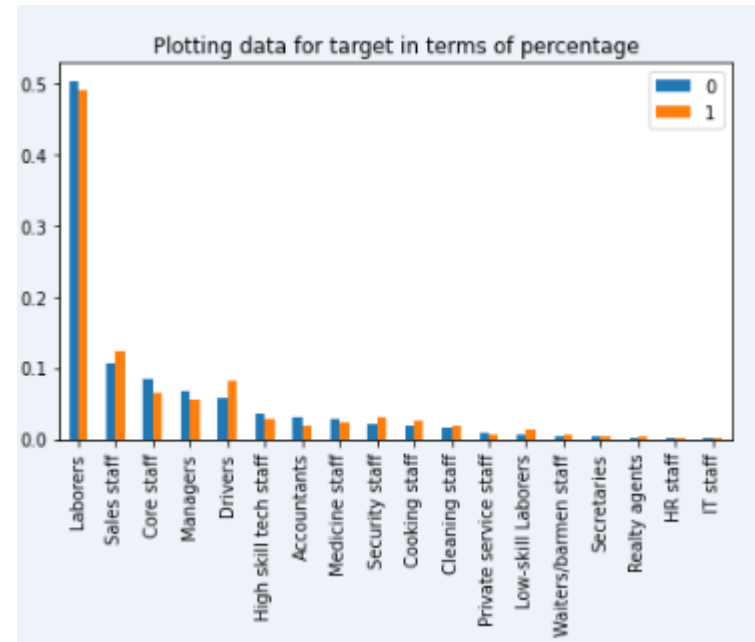
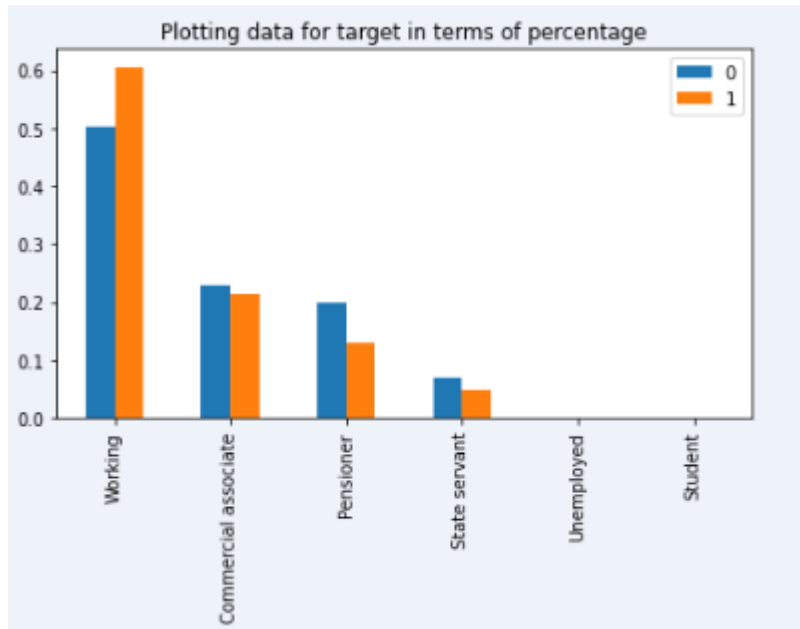
(Name\_Product\_Type)

- Around 9% of customers are walking in for loans and percentage wise they tend more towards default.
- 27% customers are being cross sold the loan products but the percentage of loan being cross sold to default prone customers are nearabout same as to potential customers.
- Company must identify the customers from previous loan history and then cross sell and should cross sell to Potential customers more.



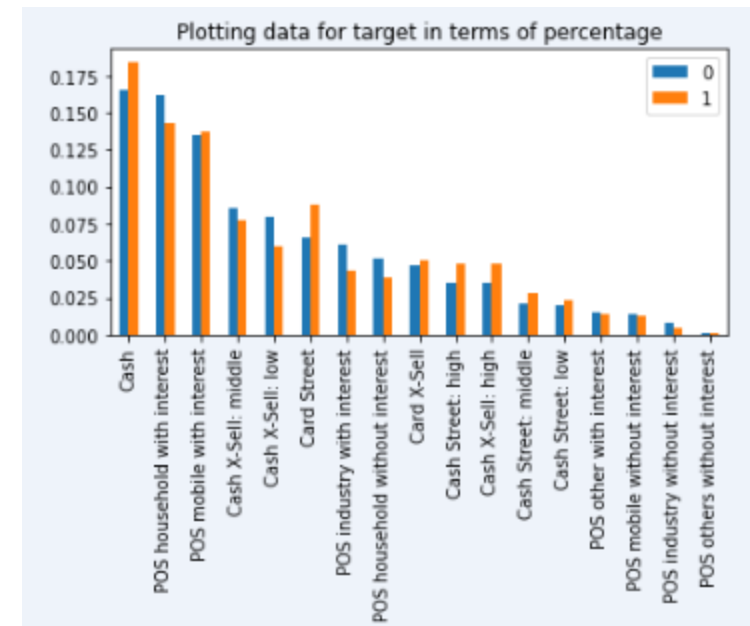
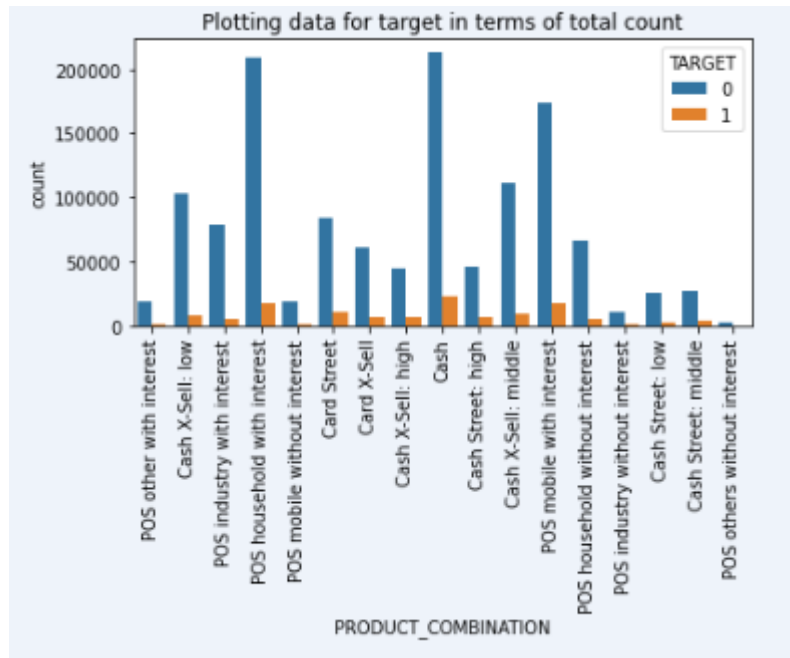
# UNIVARIATE ANALYSIS

## (INCOME TYPE AND ORGANIZATION TYPE)



- Here we can observe that customers who are of working type prone to default more.
- From other graph we can observe mostly working class people are from unorganized sector with irregular income source like (labourers, drivers etc..)
- The customers having regular income like pensioners and state servant are potential customers for company to give loans as their rate of default is lower.

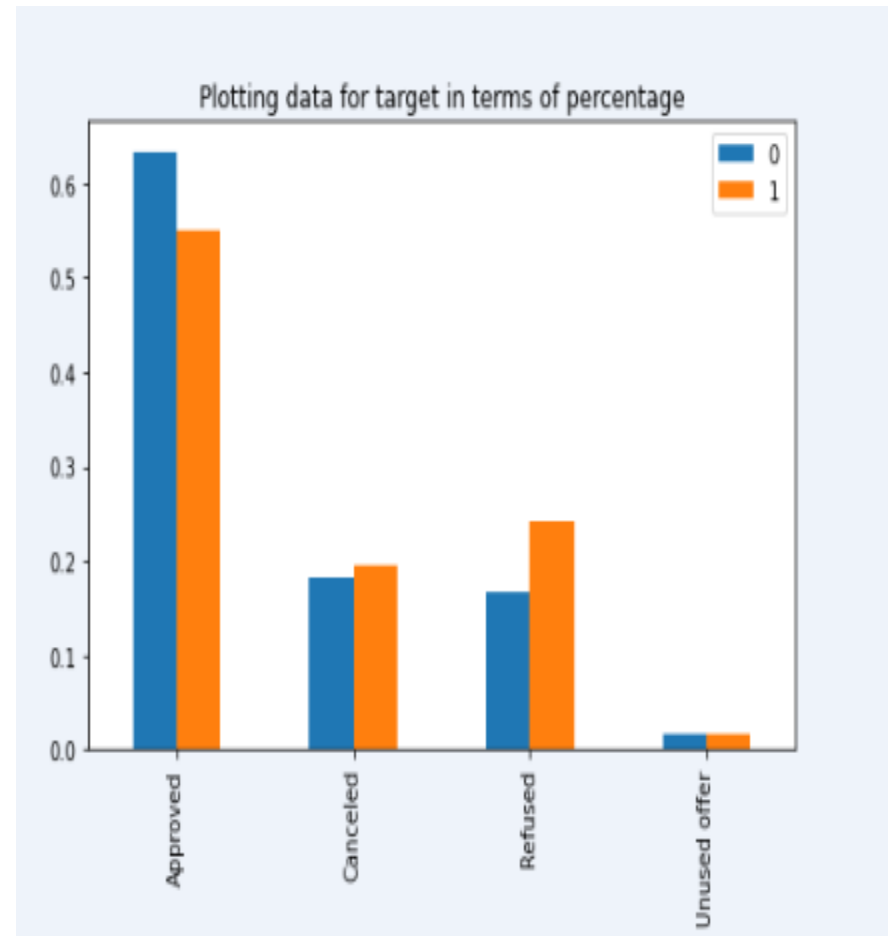
# ANALYSIS ON PRODUCT COMBINATION



Most of the loans are given in product combination of Cash and POS household with interest but cash loans are more to default while POS Household with interest loans are comparatively better as it fetches more good customers.

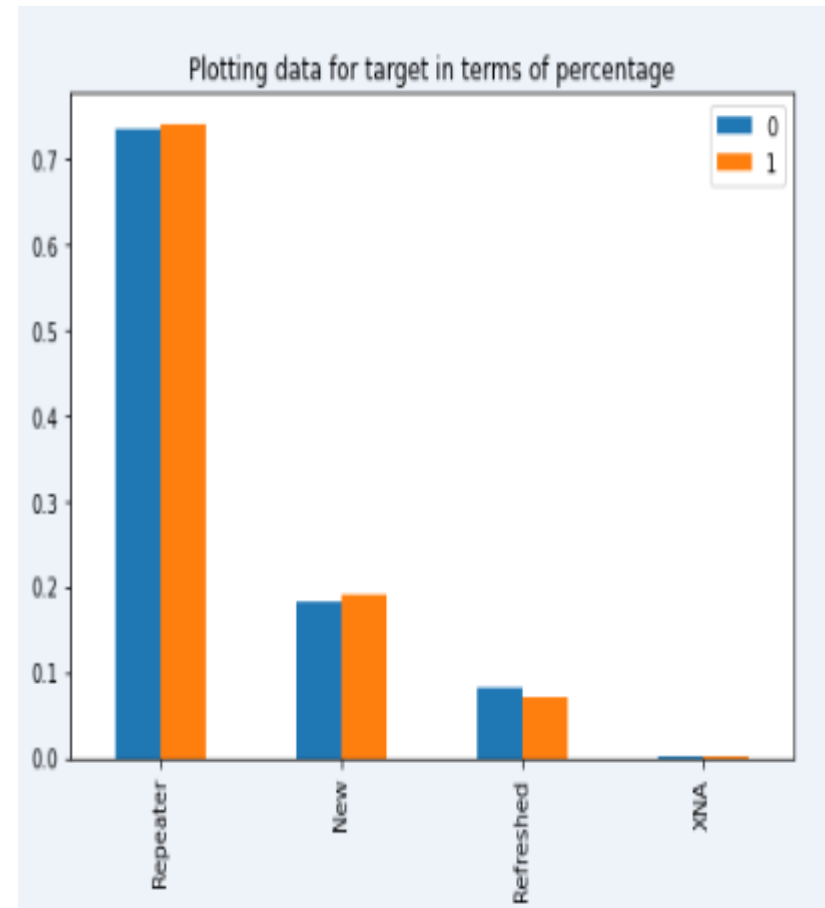
# ANALYSIS ON LOAN STATUS

- Company is refusing around 15% of loans to good customers which is leading to loss of profit.
- While the number of approved loans to default prone customers is more than 50%.
- Company should check the client background from previous credit history.



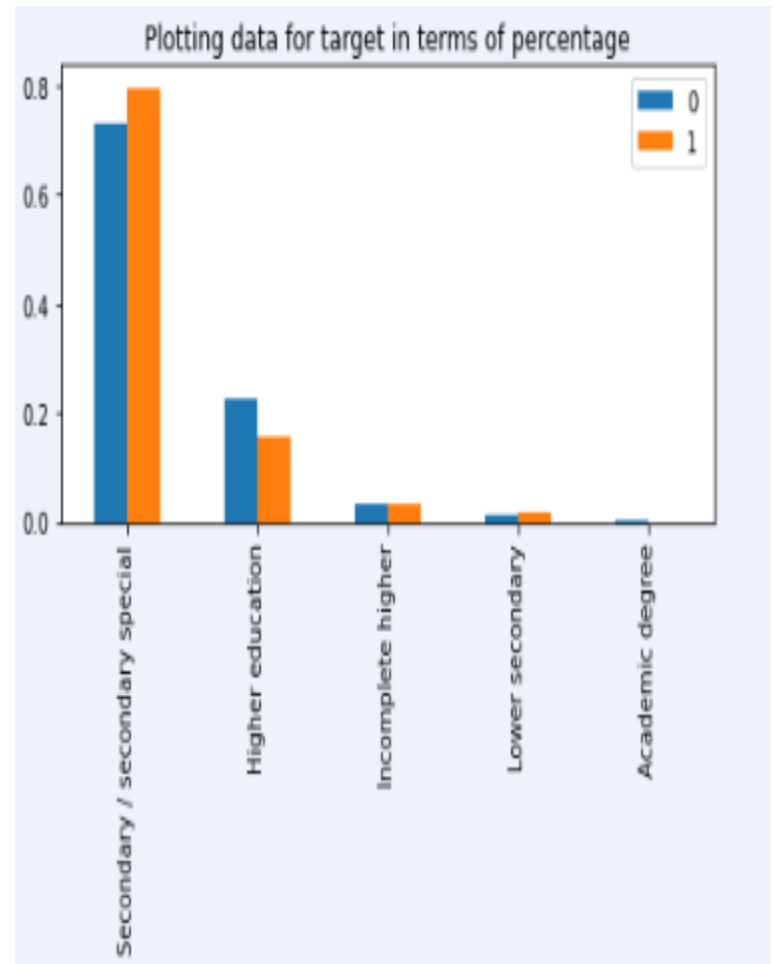
# ANALYSIS ON CLIENT TYPE

- Here again we can see that the loan given to repeater customers are prone towards default more than newer customers.
- The company should verify the repeated customers more carefully as they are already maintaining credit history with the company.



# ANALYSIS ON EDUCATION TYPE

- 80 % of Customers who have received secondary/secondary special Education are default prone.
- Customers who received higher education are our potential customers as most are paying on time.

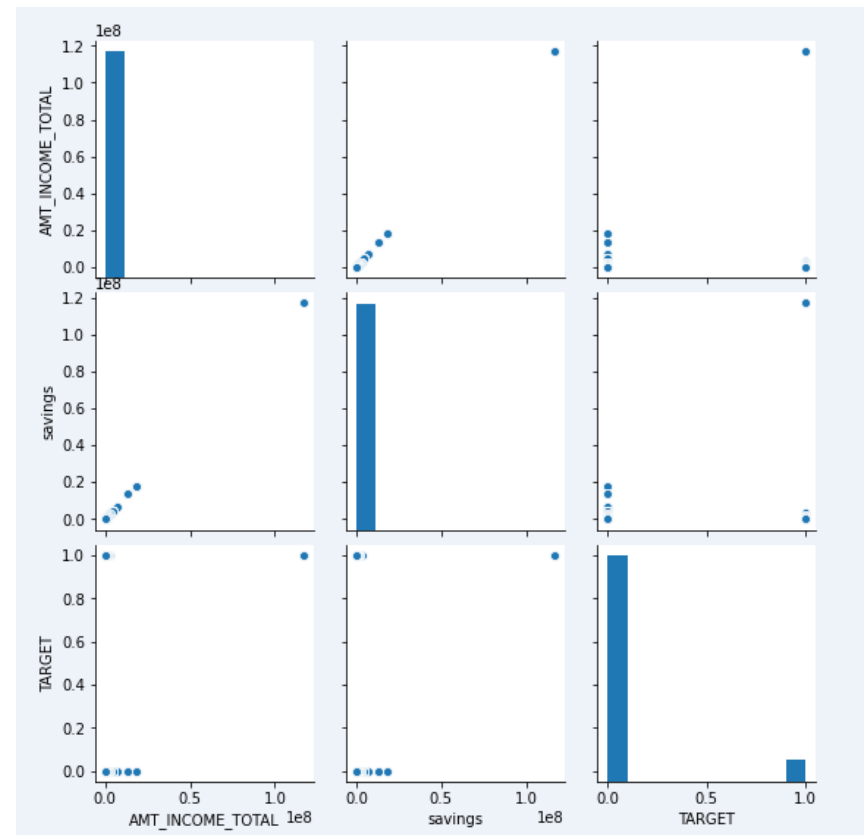




# BIVARIATE ANALYSIS

- Here we have introduced a new business driven variable called savings i.e. (Income – Annuity)
- When we analyse it against target variable it says the the customers with lower savings tend to default more.

Note- There is a outlier in amt\_credit which is to be ignored for current analysis



# TOP 10 CORRELATION

- For TARGET variable 0

- For TARGET variable 1

DAYS_EMPLOYED	FLAG_EMP_PHONE	0.999774
FLAG_EMP_PHONE	DAYS_EMPLOYED	0.999774
DAYS_BIRTH	Age_years	0.999709
Age_years	DAYS_BIRTH	0.999709
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998577
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998577
AMT_CREDIT_x	AMT_GOODS_PRICE_x	0.986271
AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.986271
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.944355
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.944355
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878468
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878468
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.875760
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.875760
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.862913
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.862913
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.835639
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.835639
AMT_ANNUITY_x	AMT_GOODS_PRICE_x	0.763257
AMT_GOODS_PRICE_x	AMT_ANNUITY_x	0.763257

FLAG_EMP_PHONE	DAYS_EMPLOYED	0.999788
DAYS_EMPLOYED	FLAG_EMP_PHONE	0.999788
DAYS_BIRTH	Age_years	0.999684
Age_years	DAYS_BIRTH	0.999684
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998378
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998378
AMT_CREDIT_x	AMT_GOODS_PRICE_x	0.982356
AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.982356
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956483
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956483
CNT_CHILDREN	CNT_FAM_MEMBERS	0.886300
CNT_FAM_MEMBERS	CNT_CHILDREN	0.886300
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.873130
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.873130
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.858283
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.858283
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.792897
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.792897
AMT_CREDIT_x	AMT_ANNUITY_x	0.745679
AMT_ANNUITY_x	AMT_CREDIT_x	0.745679

**THANK YOU**