Part 1 with Previous Applicants

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
import sklearn.preprocessing
import sklearn.model_selection
import sklearn.metrics
import sklearn.ensemble
import xgboost
```

```
In [1]: %load_ext autoreload
        %autoreload 2
        %matplotlib inline
```

```
In [2]: from home_credit_utils.utils import *
        from home_credit_utils.pd_utils import *
        from home_credit_utils.sklearn_utils import *
        from home_credit_utils.plot_utils import *
```

```
In [3]: pd.set_option('display.max_columns', 200)
        pd.set_option('display.max_rows', 200)
        pd.set_option('display.precision', 3)
        pd.set_option('display.width', 120)
        pd.set_option('display.max_colwidth', 200)
        pd.set_option('display.expand_frame_repr', False)
        pd.set_option('display.chop_threshold', 0.0001)
        pd.set_option('display.float_format', lambda x: '%.3f' % x)
        np.set_printoptions(threshold=np.inf, linewidth=200, suppress=True)
```

By Default I will show only 25 columns, there are lots more than 25 columns, this function is used:

```
In [4]: pd.set_option('display.max_columns', 25)
        pd.set_option('display.max_rows', 25)
        pd.set_option('display.precision', 2)
        pd.set_option('display.width', 80)
        pd.set_option('display.max_colwidth', 20)
        pd.set_option('display.expand_frame_repr', False)
        pd.set_option('display.chop_threshold', 0.01)
        pd.set_option('display.float_format', lambda x: '%.2f' % x)
        np.set_printoptions(threshold=10, linewidth=80, suppress=True)
```

```
In [5]: %matplotlib inline
        %load_ext autoreload
        %autoreload 2
        %aimport home_credit_utils.pd_utils, home_credit_utils.plot_utils
```

```
In [6]: %ls ../input/home-credit-default-risk/*.zip
```

```
-rw-r--r-- 1 root root 37091 Jun 3 19:12 ../input/home-credit-default-risk/HOME_CREDIT_DESCRIPTION.zip
-rw-r--r-- 1 root root 9681 Jun 3 19:12 ../input/home-credit-default-risk/sample_submission.zip
-rw-r--r-- 1 root root 158172 Jun 3 19:12 ../input/home-credit-default-risk/application_test.zip
-rw-r--r-- 1 root root 1674291 Jun 3 19:12 ../input/home-credit-default-risk/application_train.zip
```

```
In [7]: %cd ../input/home-credit-default-risk/
```

```
/kaggle/input/home-credit-default-risk
```

```
In [8]: %ls *.csv
```

```
application_train.csv  HOME_CREDIT_DESCRIPTION.csv  sample_submission.csv
application_test.csv   HOME_CREDIT_DESCRIPTION.zip  application_train.zip
application_test.zip   sample_submission.zip
```

```
In [9]: %ls -lh *.zip
```

```
total 1.6M
-rw-r--r-- 1 root root 158K Jun 3 19:12 application_test.zip
-rw-r--r-- 1 root root 1.6M Jun 3 19:12 application_train.zip
-rw-r--r-- 1 root root 37K Jun 3 19:12 HOME_CREDIT_DESCRIPTION.zip
-rw-r--r-- 1 root root 9.6K Jun 3 19:12 sample_submission.zip
```

```
In [10]: !unzip -l application_train.zip
```

```
Archive:  application_train.zip
  Length      Date    Time    Name
---------  ---------- -----   ----
  1674291  03-03-2018 07:30   application_train.csv
---------                     -------
  1674291                     1 file
```

```
In [11]: !unzip -l application_test.zip
```

```
Archive:  application_test.zip
  Length      Date    Time    Name
---------  ---------- -----   ----
   158172  03-03-2018 07:30   application_test.csv
---------                     -------
   158172                     1 file
```

```
In [12]: !unzip -l HOME_CREDIT_DESCRIPTION.zip
```

```
Archive:  HOME_CREDIT_DESCRIPTION.zip
  Length      Date    Time    Name
---------  ---------- -----   ----
    37091  03-03-2018 07:30   HOME_CREDIT_DESCRIPTION.csv
---------                     -------
    37091                     1 file
```

```
In [13]: !head -n 2 HOME_CREDIT_DESCRIPTION.csv
```

```
,Table,Row,Description
0,application_train/test,AMT_ANNUITY,For credits with a loan annuity it contains the monthly amount due to be repaid by the client.
```

```
In [14]: !head -n 2 application_train.csv
```

```
SK_ID_CURR,TARGET,NAME_CONTRACT_TYPE,CODE_GENDER,FLAG_OWN_CAR,FLAG_OWN_REALTY,CNT_CHILDREN,AMT_INCOME_TOTAL,AMT_CREDIT,AMT_ANNUITY,AMT_GOODS_PRICE,NAME_TYPE_SUITE,NAME_INCOME_TYPE,NAME_EDUCATION_TYPE,NAME_FAMILY_STATUS,NAME_HOUSING_TYPE,REGION_POPULATION_RELATIVE,DAYS_BIRTH,DAYS_EMPLOYED,DAYS_REGISTRATION,DAYS_ID_PUBLISH,OWN_CAR_AGE,FLAG_MOBIL,FLAG_EMP_PHONE,FLAG_WORK_PHONE,FLAG_CONT_MOBILE,FLAG_PHONE,FLAG_EMAIL,OCCUPATION_TYPE,CNT_FAM_MEMBERS,REGION_RATING_CITY,WEEKDAY_APPR_PROCESS_START,HOUR_APPR_PROCESS_START,REG_REGION_NOT_LIVE_REGION,REG_REGION_NOT_WORK_REGION,LIVE_REGION_NOT_WORK_REGION,REG_CITY_NOT_LIVE_CITY,REG_CITY_NOT_WORK_CITY,LIVE_CITY_NOT_WORK_CITY,ORGANIZATION_TYPE,EXT_SOURCE_1,EXT_SOURCE_2,EXT_SOURCE_3,APARTMENTS_AVG,BASEMENTAREA_AVG,YEARS_BEGINEXPLUATATION_AVG,YEARS_COMMON,ELEVATORS_AVG,ENTRANCES_AVG,NONLIVINGAREA_AVG,NONLIVINGAPARTMENTS_AVG,LIVINGAPARTMENTS_AVG,LIVINGAREA_AVG,APARTMENTS_MODE,BASEMENTAREA_MODE,YEARS_BEGINEXPLUATATION_MODE,YEARS_COMMON,ELEVATORS_MODE,ENTRANCES_MODE,NONLIVINGAREA_MODE,NONLIVINGAPARTMENTS_MODE,LIVINGAPARTMENTS_MODE,LIVINGAREA_MODE,APARTMENTS_MEDI,BASEMENTAREA_MEDI,YEARS_BEGINEXPLUATATION_MEDI,YEARS_COMMON,ELEVATORS_MEDI,ENTRANCES_MEDI,NONLIVINGAREA_MEDI,NONLIVINGAPARTMENTS_MEDI,LIVINGAPARTMENTS_MEDI,LIVINGAREA_MEDI,FONDKAPREMONT_MODE,HOUSETYPE_MODE,TOTALAREA_MODE,WALLSMATERIAL_MODE,EMERGENCYSTATE_MODE,DEF_30_CNT_SOCIAL_CIRCLE,DEF_30_CNT_SOCIAL_CIRCLE,DEF_60_CNT_SOCIAL_CIRCLE,DEF_60_CNT_SOCIAL_CIRCLE,DAYS_LAST_PHONE_CHANGE,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,AMT_REQ_CREDIT_BUREAU_HOUR,AMT_REQ_CREDIT_BUREAU_DAY,AMT_REQ_CREDIT_BUREAU_WEEK,AMT_REQ_CREDIT_BUREAU_MON,AMT_REQ_CREDIT_BUREAU_QRT,AMT_REQ_CREDIT_BUREAU_YEAR
```

```
In [15]: %ls -lh *.zip
```

```
total 1.6M
-rw-r--r-- 1 root root 158K Jun 3 19:12 application_test.zip
-rw-r--r-- 1 root root 1.6M Jun 3 19:12 application_train.zip
-rw-r--r-- 1 root root 37K Jun 3 19:12 HOME_CREDIT_DESCRIPTION.zip
-rw-r--r-- 1 root root 9.6K Jun 3 19:12 sample_submission.zip
```

```
In [16]: %ls -lh *.csv
```

```
-rw-r--r-- 1 root root 37K Jun 3 19:12 HOME_CREDIT_DESCRIPTION.csv
-rw-r--r-- 1 root root 9.6K Jun 3 19:12 sample_submission.csv
-rw-r--r-- 1 root root 158K Jun 3 19:12 application_test.csv
-rw-r--r-- 1 root root 1.6M Jun 3 19:12 application_train.csv
```

```
In [17]: %ls -lh ../*/*/*
```

```
-rw-r--r-- 1 root root 1.6K Jun 3 19:12 ../input/home-credit-default-risk/kernel-metadata.txt
-rw-r--r-- 1 root root 1.6K Jun 3 19:12 ../input/home-credit-default-risk/kernel-metadata.bak
```

```
In [18]: %ls -lh ../input/home-credit-default-risk/*.zip
```

```
-rw-r--r-- 1 root root 158K Jun 3 19:12 ../input/home-credit-default-risk/application_test.zip
-rw-r--r-- 1 root root 1.6M Jun 3 19:12 ../input/home-credit-default-risk/application_train.zip
-rw-r--r-- 1 root root 37K Jun 3 19:12 ../input/home-credit-default-risk/HOME_CREDIT_DESCRIPTION.zip
-rw-r--r-- 1 root root 9.6K Jun 3 19:12 ../input/home-credit-default-risk/sample_submission.zip
```

```
In [19]: %ls -lh ../input/home-credit-default-risk/*.csv
```

```
-rw-r--r-- 1 root root 37K Jun 3 19:12 ../input/home-credit-default-risk/HOME_CREDIT_DESCRIPTION.csv
-rw-r--r-- 1 root root 9.6K Jun 3 19:12 ../input/home-credit-default-risk/sample_submission.csv
-rw-r--r-- 1 root root 158K Jun 3 19:12 ../input/home-credit-default-risk/application_test.csv
-rw-r--r-- 1 root root 1.6M Jun 3 19:12 ../input/home-credit-default-risk/application_train.csv
```

```
In [20]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [21]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [22]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [23]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [24]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [25]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [26]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [27]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [28]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [29]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [30]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [31]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [32]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [33]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [34]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [35]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [36]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [37]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [38]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [39]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [40]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [41]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [42]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [43]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [44]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [45]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [46]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [47]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [48]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [49]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [50]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [51]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [52]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [53]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [54]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [55]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [56]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [57]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [58]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [59]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [60]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [61]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [62]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [63]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [64]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [65]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [66]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [67]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [68]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [69]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [70]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [71]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [72]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [73]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [74]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [75]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [76]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [77]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [78]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [79]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [80]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [81]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [82]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [83]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [84]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [85]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [86]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [87]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [88]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [89]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [90]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [91]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [92]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [93]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [94]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [95]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [96]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [97]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [98]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [99]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [100]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [101]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [102]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [103]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [104]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [105]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [106]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [107]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [108]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [109]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [110]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [111]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [112]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [113]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [114]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [115]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [116]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [117]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [118]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [119]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [120]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [121]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [122]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [123]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [124]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [125]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [126]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [127]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [128]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [129]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [130]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [131]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [132]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [133]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [134]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [135]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [136]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [137]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [138]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [139]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [140]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [141]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [142]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [143]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [144]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [145]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [146]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [147]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [148]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [149]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [150]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [151]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [152]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [153]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [154]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [155]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [156]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [157]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [158]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [159]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [160]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [161]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [162]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [163]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [164]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [165]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [166]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [167]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [168]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [169]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [170]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [171]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [172]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [173]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [174]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [175]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [176]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [177]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [178]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [179]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [180]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [181]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [182]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [183]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [184]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [185]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [186]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [187]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [188]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [189]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [190]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [191]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [192]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [193]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [194]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [195]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [196]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [197]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [198]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [199]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [200]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [201]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [202]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [203]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [204]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [205]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [206]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [207]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [208]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [209]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [210]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [211]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [212]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [213]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [214]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [215]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [216]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [217]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [218]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [219]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [220]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [221]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [222]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [223]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [224]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [225]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [226]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [227]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [228]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [229]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [230]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [231]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [232]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [233]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [234]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [235]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [236]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [237]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [238]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [239]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [240]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [241]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [242]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [243]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [244]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [245]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [246]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [247]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [248]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [249]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [250]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [251]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [252]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [253]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [254]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [255]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [256]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [257]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [258]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [259]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [260]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [261]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [262]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [263]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [264]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [265]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [266]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [267]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [268]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -c
```

```
1
```

```
In [269]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -c
```

```
1
```

```
In [270]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -l
```

```
4
```

```
In [271]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -l
```

```
4
```

```
In [272]: %ls -lh ../input/home-credit-default-risk/*.zip | wc -w
```

```
1
```

```
In [273]: %ls -lh ../input/home-credit-default-risk/*.csv | wc -w
```

```
1
```

```
In [274]: %ls -lh ../input/home-credit-default-ris
```

Calculating Standard deviations to check values fall outside SD

```
Cmd + Shift + Enter, ctrl + shift + enter
DAYS_BIRTH_sd = 4.202
DAYS_EMPLOYED_sd = 143.8506
DAYS_REGISTRATION_sd = 4.202
DAYS_ID_PUBLISH_sd = 2.7315
CNT_CHILDREN_sd = 0.7127
CNT_FAM_MEMBERS_sd = 0.9112
AMT_INCOME_TOTAL_sd = 237.6
AMT_CREDIT_sd = 402.5
AMT_ANNUITY_sd = 14.493
AMT_GOODS_PRICE_sd = 359.8
```

```
Cmd + Shift + Enter, ctrl + shift + enter
DAYS_BIRTH_mean = -15768.05
DAYS_EMPLOYED_mean = 67724.74
DAYS_REGISTRATION_mean = -4986.12
DAYS_ID_PUBLISH_mean = -2994.202
CNT_CHILDREN_mean = 0.4171
CNT_FAM_MEMBERS_mean = 2.153
AMT_INCOME_TOTAL_mean = 168797.9
AMT_CREDIT_mean = 599026
AMT_ANNUITY_mean = 27108.57
AMT_GOODS_PRICE_mean = 538396.2
```

In [24]: plt.rc('font',family='Arial')
sns.scatterplot(data=df,x='AMT_CREDIT',y='AMT_GOODS_PRICE',s=5,hue='CODE_GENDER',palette=['red','yellow','green','blue'])
plt.xticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'],rotation=90)
plt.yticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'])
plt.title('Relationship between AMT_CREDIT and AMT_GOODS_PRICE',size=15,weight='bold')
plt.xlabel('AMT_CREDIT',size=15,weight='bold')
plt.ylabel('AMT_GOODS_PRICE',size=15,weight='bold')
plt.legend(title='CODE_GENDER',loc='upper left',bbox_to_anchor=(1,1))
plt.show()

In [25]: plt.rc('font',family='Arial')
sns.scatterplot(data=df,x='AMT_CREDIT',y='AMT_GOODS_PRICE',s=5,hue='FLAG_OWN_CAR',palette=['red','yellow','green','blue'])
plt.xticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'],rotation=90)
plt.yticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'])
plt.title('Relationship between AMT_CREDIT and AMT_GOODS_PRICE',size=15,weight='bold')
plt.xlabel('AMT_CREDIT',size=15,weight='bold')
plt.ylabel('AMT_GOODS_PRICE',size=15,weight='bold')
plt.legend(title='FLAG_OWN_CAR',loc='upper left',bbox_to_anchor=(1,1))
plt.show()

In [26]: plt.rc('font',family='Arial')
sns.scatterplot(data=df,x='AMT_CREDIT',y='AMT_GOODS_PRICE',s=5,hue='FLAG_OWN_REALTY',palette=['red','yellow','green','blue'])
plt.xticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'],rotation=90)
plt.yticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'])
plt.title('Relationship between AMT_CREDIT and AMT_GOODS_PRICE',size=15,weight='bold')
plt.xlabel('AMT_CREDIT',size=15,weight='bold')
plt.ylabel('AMT_GOODS_PRICE',size=15,weight='bold')
plt.legend(title='FLAG_OWN_REALTY',loc='upper left',bbox_to_anchor=(1,1))
plt.show()

In [27]: plt.rc('font',family='Arial')
sns.scatterplot(data=df,x='AMT_CREDIT',y='AMT_GOODS_PRICE',s=5,hue='NAME_EDUCATION_TYPE',palette=['red','yellow','green','blue'])
plt.xticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'],rotation=90)
plt.yticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'])
plt.title('Relationship between AMT_CREDIT and AMT_GOODS_PRICE',size=15,weight='bold')
plt.xlabel('AMT_CREDIT',size=15,weight='bold')
plt.ylabel('AMT_GOODS_PRICE',size=15,weight='bold')
plt.legend(title='NAME_EDUCATION_TYPE',loc='upper left',bbox_to_anchor=(1,1))
plt.show()

In [28]: plt.rc('font',family='Arial')
sns.scatterplot(data=df,x='AMT_CREDIT',y='AMT_GOODS_PRICE',s=5,hue='NAME_INCOME_TYPE',palette=['red','yellow','green','blue'])
plt.xticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'],rotation=90)
plt.yticks(ticks=[0,100000,200000,300000,400000,500000,600000,700000,800000,900000],labels=['0','100K','200K','300K','400K','500K','600K','700K','800K','900K'])
plt.title('Relationship between AMT_CREDIT and AMT_GOODS_PRICE',size=15,weight='bold')
plt.xlabel('AMT_CREDIT',size=15,weight='bold')
plt.ylabel('AMT_GOODS_PRICE',size=15,weight='bold')
plt.legend(title='NAME_INCOME_TYPE',loc='upper left',bbox_to_anchor=(1,1))
plt.show()

Conclusion:

- All were Defaulters out of 1 Million Applications.

Gender Distribution:

- Female Applicants are 70% - 75% after unique online clients that are particular meaning to 5 But our Analysis Continued with those would answer.
- Male Applicants are Thirty one times high 25% were as minimum which are less equal with 1% Highlight which is extreme all remaining Gender column will remove the record would be our analysis.
- Female have relatively debt of debts 20% as the debts are dynamic they were removed from our analysis.

Finding:

- Highest Applicants were from Consumer loans & Cash loans but Defaulters were quite high from Cash loan applicants.
- Highest Ms with applicants almost least applicants were Considered Refused Loan.
- Highest Applicants were from Working Income type but Defaulters were quite high from Commercial Income type & Million are purely should be given to the customers who relocate Office more than 2 times.
- Highest Applicants were from Secondary / secondary special Education type but Defaulters is Million.
- Most of the applicants Housing Situation lives within the age of 1 year, which is a risky factor.
- Highest Applicants were from Single / not married Family Status but Defaulters were quite high from Civil and Married.
- Defaulters were high in Cash POS Mobile with Interest POS Installed with Interest and Cash Street.
- Highest Applicants were from Cash loans but Defaulters were quite high from Revolving loans.
- Highest Defaulters were in Consumer requested applications where Bad not Informed.
- Highest Defaulters were in Cash loans requested applications where Bad not Informed.

In [29]: from zipfile import zipfile,ZIP_DEFLATED,ZipInfo
zf = zipfile.ZipFile('Home_Credit.zip', mode='w', compression=ZIP_DEFLATED)
for c in [*]: