NAME-AKSHAY ANAND

PROJECT DESCRIPTION – This project is about a case study relayed to a Bank Loan. We have to carry out an EDA (Exploratory Data Analysis).

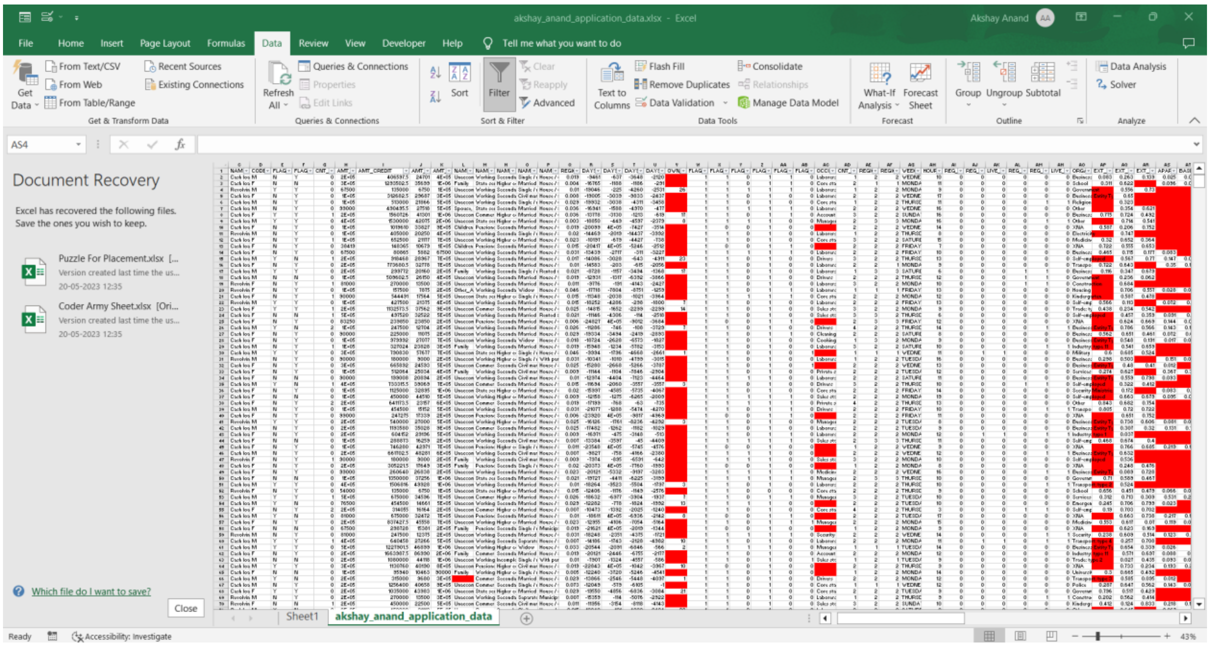
- Present the overall approach of the **analysis**. Mention the problem statement and the analysis approach briefly

SINCE THE DATASET HAD A LOT OF MISSING VALUES WE FILLED THE BLANK VALUES WITH MEAN OR MEDIAN DEPENDING ON OUTLIERS AND THEN USED PIVOT TABLES,FILTERS AND OTHER FEATURES OF EXCEL TO PERFORM FURTHER ANALYSIS.

TECH STACK USED: MS EXCEL

- **Indentify** the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)
  Hint: *Note that in EDA, since it is not necessary to replace the missing value, but if you have to replace the missing value, what should be the approach. Clearly mention the approach.*

TO FIND OUT BLANK OR MISSING VALUES WE CAN USE FILTERS AND SELECT BLANK OR USING FIND & REPLACE WE REPLACE ALL BLANK CELLS BY ANY COLOUR.



Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

Puzzle For Placement.xlsx [...
Version created last time the us...
20-05-2023 12:35

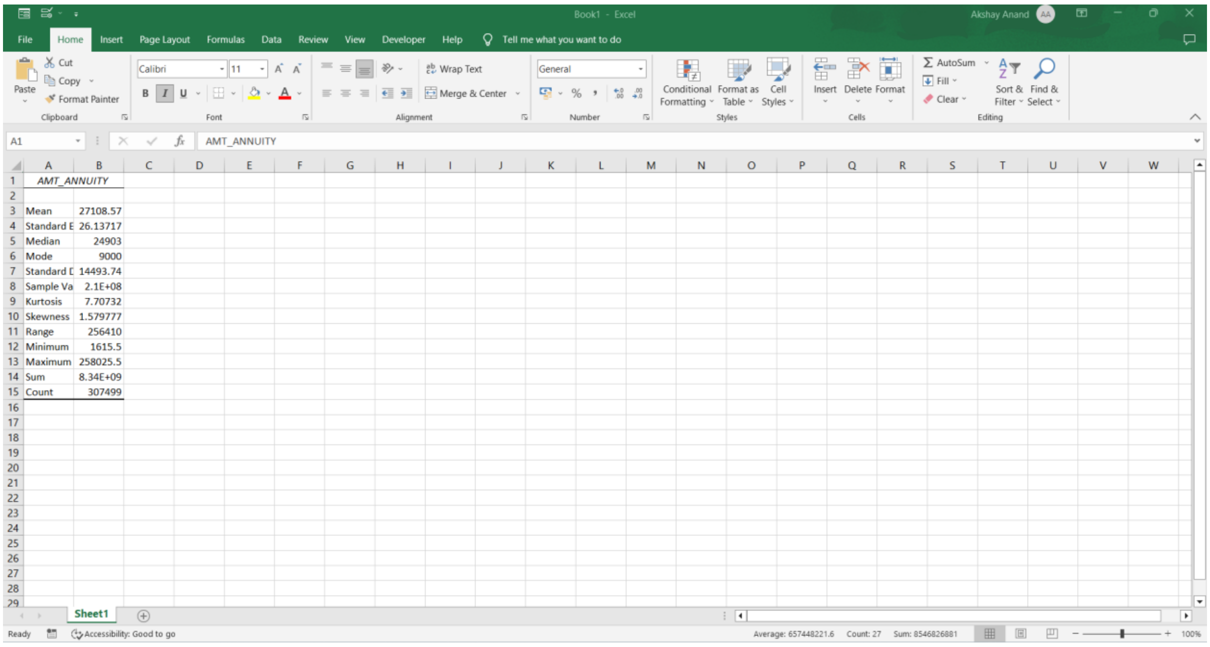
Coder Army Sheet.xlsx [Ori...
Version created last time the us...
20-05-2023 12:35

Which file do I want to save?

Close

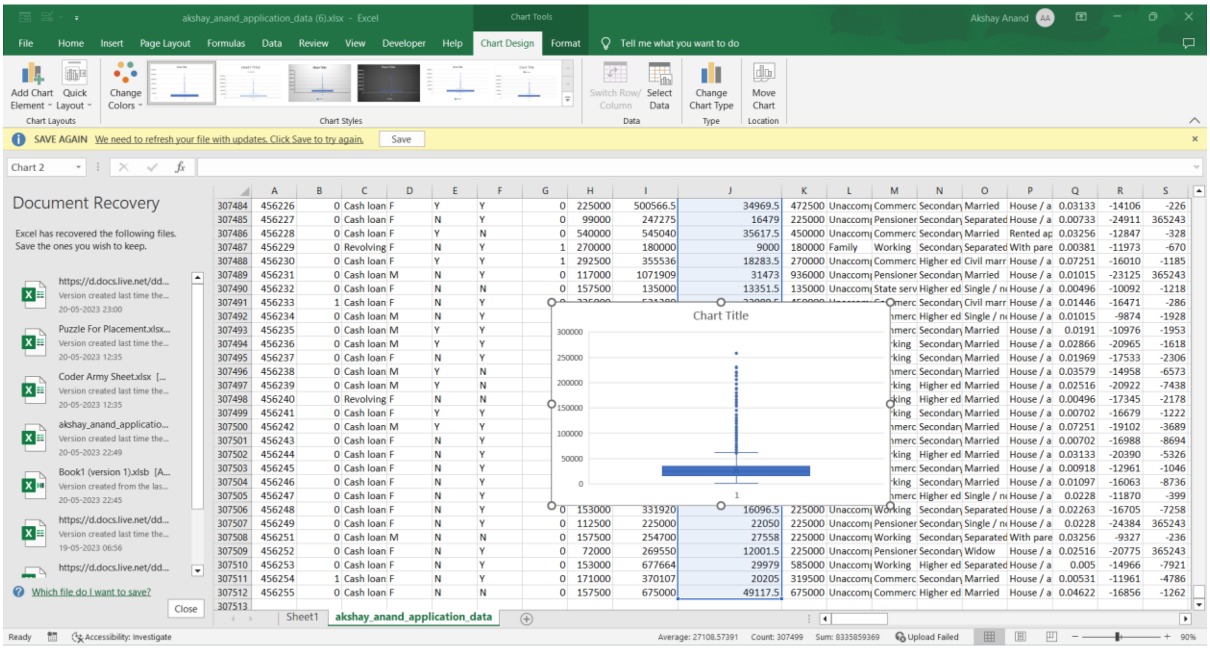
WE CAN USE DESCRIPTIVE STATISTICS BY SELECTING DATA AND THEN DATA ANALYSIS.

SINCE IT IS POSSIBLE THAT DATA HAS OUTLIERS SO WE REPLACE ALL MISSING VALUES BY MEDIAN.



	A	B
1	*AMT_ANNUITY*	
2		
3	Mean	27108.57
4	Standard E	26.13717
5	Median	24903
6	Mode	9000
7	Standard D	14493.74
8	Sample Va	2.1E+08
9	Kurtosis	7.70732
10	Skewness	1.579777
11	Range	256410
12	Minimum	1615.5
13	Maximum	258025.5
14	Sum	8.34E+09
15	Count	307499

SINCE THE MEDIAN FOR AMT_ANNUITY COLUMN IS 24903 WE REPLACE ALL BLANK VALUES IN THAT COLUMN WITH THIS MEDIAN VALUE.



SAVE AGAIN We need to refresh your file with updates. Click Save to try again. Save

Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

https://d.docs.live.net/dd... Version created last time the... 20-05-2023 23:00

Puzzle For Placement.xlsx... Version created last time the... 20-05-2023 12:35

Coder Army Sheet.xlsx [... Version created last time the... 20-05-2023 12:35

akshay_anand_applicatio... Version created last time the... 20-05-2023 22:49

Book1 (version 1).xlsb [A... Version created from the las... 20-05-2023 22:45

https://d.docs.live.net/dd... Version created last time the... 19-05-2023 06:56

https://d.docs.live.net/dd...

Which file do I want to save?

Close

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
307484	456226	0	Cash loan	F	Y	Y	0	225000	500566.5	34969.5	472500	Unaccom	Commerc	Secondary	Married	House / a	0.03133	-14106	-226
307485	456227	0	Cash loan	F	N	Y	0	99000	247275	16479	225000	Unaccom	Pensioner	Secondary	Separated	House / a	0.00733	-24911	365243
307486	456228	0	Cash loan	F	Y	N	0	540000	545040	35617.5	450000	Unaccom	Commerc	Secondary	Married	Rented ap	0.03256	-12847	-328
307487	456229	0	Revolving	F	N	Y	1	270000	180000	9000	180000	Family	Working	Secondary	Separated	With pare	0.00381	-11973	-670
307488	456230	0	Cash loan	F	Y	Y	1	292500	355536	18283.5	270000	Unaccom	Commerc	Higher ed	Civil marr	House / a	0.07251	-16010	-1185
307489	456231	0	Cash loan	M	N	Y	0	117000	1071909	31473	936000	Unaccom	Pensioner	Secondary	Married	House / a	0.01015	-23125	365243
307490	456232	0	Cash loan	F	N	N	0	157500	135000	13351.5	135000	Unaccom	State serv	Higher ed	Single / n	House / a	0.00496	-10092	-1218
307491	456233	1	Cash loan	F	N	Y							merc	Secondary	Civil marr	House / a	0.01446	-16471	-286
307492	456234	0	Cash loan	M	N	Y							merc	Higher ed	Single / n	House / a	0.01015	-9874	-1928
307493	456235	0	Cash loan	M	Y	Y							merc	Secondary	Married	House / a	0.0191	-10976	-1953
307494	456236	0	Cash loan	M	Y	Y							king	Secondary	Married	House / a	0.02866	-20965	-1618
307495	456237	0	Cash loan	F	N	Y							king	Secondary	Married	House / a	0.01969	-17533	-2306
307496	456238	0	Cash loan	M	Y	N							merc	Secondary	Married	House / a	0.03579	-14958	-6573
307497	456239	0	Cash loan	M	Y	N							king	Higher ed	Married	House / a	0.02516	-20922	-7438
307498	456240	0	Revolving	F	N	N							king	Higher ed	Married	House / a	0.00496	-17345	-2178
307499	456241	0	Cash loan	F	Y	Y							king	Secondary	Married	House / a	0.00702	-16679	-1222
307500	456242	0	Cash loan	M	Y	Y							merc	Secondary	Married	House / a	0.07251	-19102	-3689
307501	456243	0	Cash loan	F	N	Y							merc	Secondary	Married	House / a	0.00702	-16988	-8694
307502	456244	0	Cash loan	F	N	Y							king	Higher ed	Married	House / a	0.03133	-20390	-5326
307503	456245	0	Cash loan	F	N	Y							merc	Secondary	Married	House / a	0.00918	-12961	-1046
307504	456246	0	Cash loan	F	N	Y							king	Secondary	Married	House / a	0.01097	-16063	-8736
307505	456247	0	Cash loan	F	N	Y							merc	Higher ed	Single / n	House / a	0.0228	-11870	-399
307506	456248	0	Cash loan	F	N	Y	0	153000	331920	16096.5	225000	Unaccom	Working	Secondary	Separated	House / a	0.02263	-16705	-7258
307507	456249	0	Cash loan	F	N	Y	0	112500	225000	22050	225000	Unaccom	Pensioner	Secondary	Single / n	House / a	0.0228	-24384	365243
307508	456251	0	Cash loan	M	N	N	0	157500	254700	27558	225000	Unaccom	Working	Secondary	Separated	With pare	0.03256	-9327	-236
307509	456252	0	Cash loan	F	N	Y	0	72000	269550	12001.5	225000	Unaccom	Pensioner	Secondary	Widow	House / a	0.02516	-20775	365243
307510	456253	0	Cash loan	F	N	Y	0	153000	677664	29979	585000	Unaccom	Working	Higher ed	Separated	House / a	0.005	-14966	-7921
307511	456254	1	Cash loan	F	N	Y	0	171000	370107	20205	319500	Unaccom	Commerc	Secondary	Married	House / a	0.00531	-11961	-4786
307512	456255	0	Cash loan	F	N	N	0	157500	675000	49117.5	675000	Unaccom	Commerc	Higher ed	Married	House / a	0.04622	-16856	-1262
307513																			

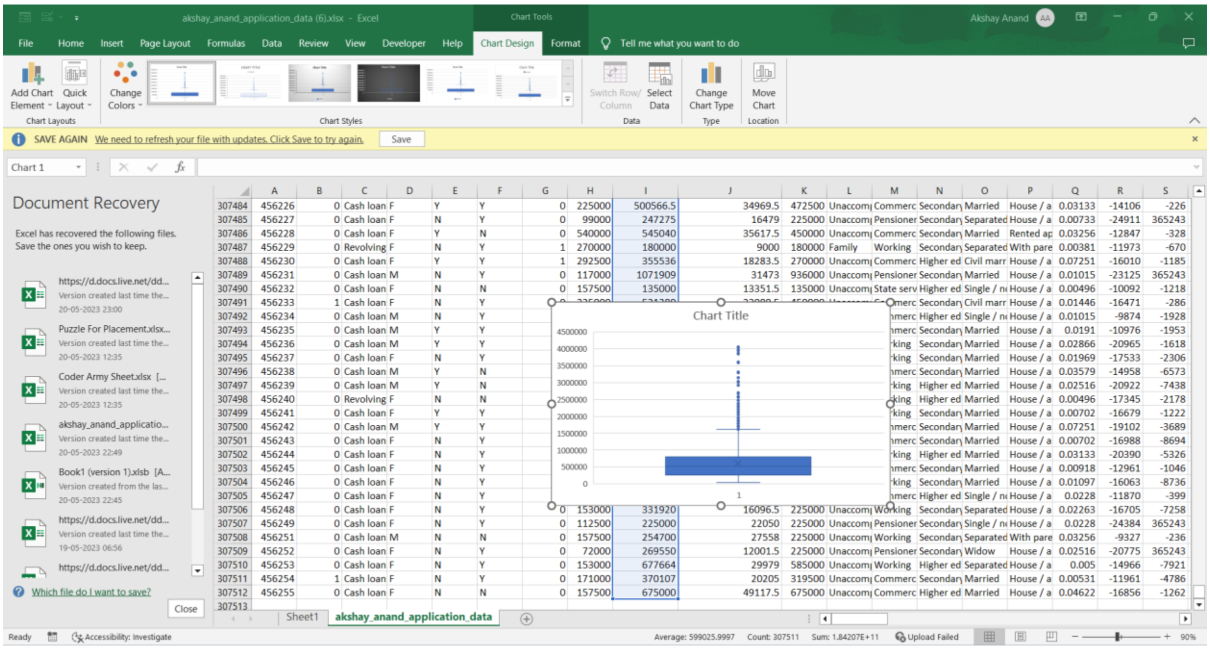
Average: 27108.57391 Count: 307499 Sum: 8335859369

- Identify if there are **outliers** in the dataset. Also, mention why do you think it is an outlier. Again, remember that for this exercise, it is not necessary to remove any data points.

THE POINTS WHICH ARE LOCATED OUTSIDE THE WHISKERS OF THE BOX PLOT ARE OUTLIERS.

ALSO, THE MEAN IS GREATLY AFFECTED BY OUTLIERS BUT MODE AND MEDIAN ARE NOT MUCH AFFECTED BY OUTLIERS SO IF ANY DATA HAS A SIGNIFICANT DIFFERENCE BETWEEN MEAN AND MEDIAN THEN WE CAN SAY THAT THERE ARE OUTLIERS IN THE DATA.

FOR AMT_CREDIT COLUMN



SAVE AGAIN We need to refresh your file with updates. Click Save to try again. Save

Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

https://d.docs.live.net/dd... Version created last time the... 20-05-2023 23:00

Puzzle For Placement.xlsx... Version created last time the... 20-05-2023 12:35

Coder Army Sheet.xlsx [... Version created last time the... 20-05-2023 12:35

akshay_anand_applicatio... Version created last time the... 20-05-2023 22:49

Book1 (version 1).xlsb [A... Version created from the las... 20-05-2023 22:45

https://d.docs.live.net/dd... Version created last time the... 19-05-2023 06:56

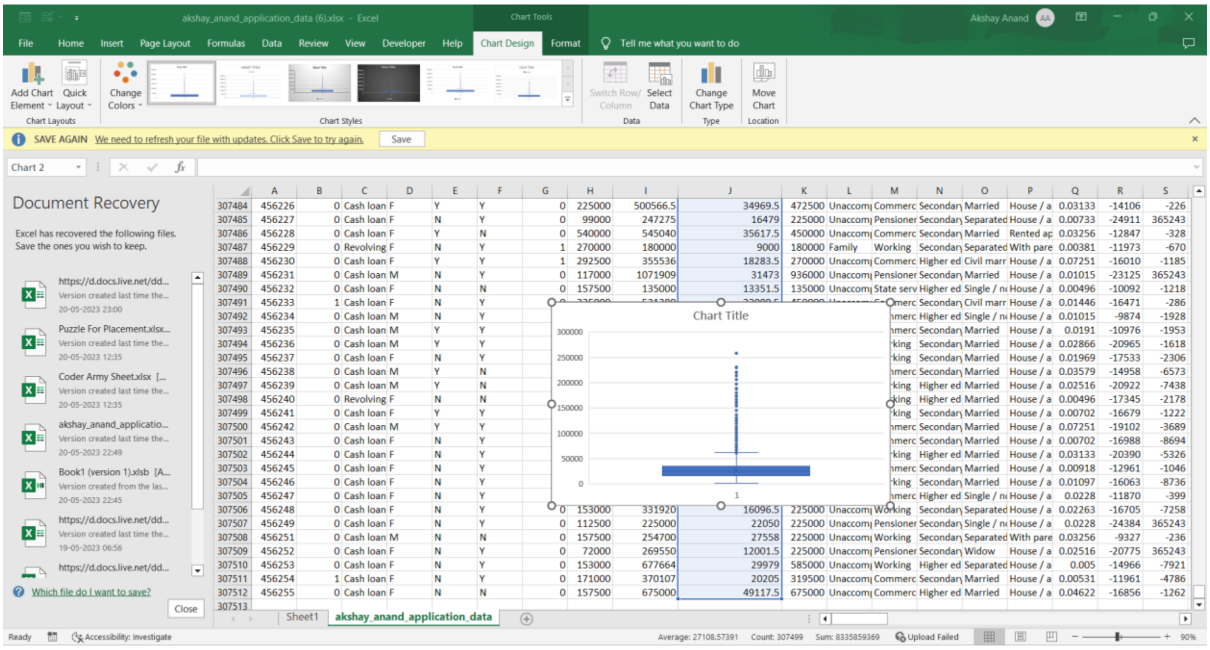
https://d.docs.live.net/dd...

Which file do I want to save?

Close

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
307484	456226	0	Cash loan	F	Y	Y	0	225000	500566.5	34969.5	472500	Unaccom	Commerc	Secondar	Married	House / a	0.03133	-14106	-226
307485	456227	0	Cash loan	F	N	Y	0	99000	247275	16479	225000	Unaccom	Pensioner	Secondar	Separated	House / a	0.00733	-24911	365243
307486	456228	0	Cash loan	F	Y	N	0	540000	545040	35617.5	450000	Unaccom	Commerc	Secondar	Married	Rented ap	0.03256	-12847	-328
307487	456229	0	Revolving	F	N	Y	1	270000	180000	9000	180000	Family	Working	Secondar	Separated	With pare	0.00381	-11973	-670
307488	456230	0	Cash loan	F	Y	Y	1	292500	355536	18283.5	270000	Unaccom	Commerc	Higher ed	Civil marr	House / a	0.07251	-16010	-1185
307489	456231	0	Cash loan	M	N	Y	0	117000	1071909	31473	936000	Unaccom	Pensioner	Secondar	Married	House / a	0.01015	-23125	365243
307490	456232	0	Cash loan	F	N	N	0	157500	135000	13351.5	135000	Unaccom	State serv	Higher ed	Single / n	House / a	0.00496	-10092	-1218
307491	456233	1	Cash loan	F	N	Y							merc	Secondar	Civil marr	House / a	0.01446	-16471	-286
307492	456234	0	Cash loan	M	N	Y							merc	Higher ed	Single / n	House / a	0.01015	-9874	-1928
307493	456235	0	Cash loan	M	Y	Y							merc	Secondar	Married	House / a	0.0191	-10976	-1953
307494	456236	0	Cash loan	M	Y	Y							king	Secondar	Married	House / a	0.02866	-20965	-1618
307495	456237	0	Cash loan	F	N	Y							king	Secondar	Married	House / a	0.01969	-17533	-2306
307496	456238	0	Cash loan	M	Y	N							merc	Secondar	Married	House / a	0.03579	-14958	-6573
307497	456239	0	Cash loan	M	Y	N							king	Higher ed	Married	House / a	0.02516	-20922	-7438
307498	456240	0	Revolving	F	N	N							king	Higher ed	Married	House / a	0.00496	-17345	-2178
307499	456241	0	Cash loan	F	Y	Y							king	Secondar	Married	House / a	0.00702	-16679	-1222
307500	456242	0	Cash loan	M	Y	Y							merc	Secondar	Married	House / a	0.07251	-19102	-3689
307501	456243	0	Cash loan	F	N	Y							merc	Secondar	Married	House / a	0.00702	-16988	-8694
307502	456244	0	Cash loan	F	N	Y							king	Higher ed	Married	House / a	0.03133	-20390	-5326
307503	456245	0	Cash loan	F	N	Y							merc	Secondar	Married	House / a	0.00918	-12961	-1046
307504	456246	0	Cash loan	F	N	Y							king	Secondar	Married	House / a	0.01097	-16063	-8736
307505	456247	0	Cash loan	F	N	Y							merc	Higher ed	Single / n	House / a	0.0228	-11870	-399
307506	456248	0	Cash loan	F	N	Y	0	153000	331920	16096.5	225000	Unaccom	Working	Secondar	Separated	House / a	0.02263	-16705	-7258
307507	456249	0	Cash loan	F	N	Y	0	112500	225000	22050	225000	Unaccom	Pensioner	Secondar	Single / n	House / a	0.0228	-24384	365243
307508	456251	0	Cash loan	M	N	N	0	157500	254700	27558	225000	Unaccom	Working	Secondar	Separated	With pare	0.03256	-9327	-236
307509	456252	0	Cash loan	F	N	Y	0	72000	269550	12001.5	225000	Unaccom	Pensioner	Secondar	Widow	House / a	0.02516	-20775	365243
307510	456253	0	Cash loan	F	N	Y	0	153000	677664	29979	585000	Unaccom	Working	Higher ed	Separated	House / a	0.005	-14966	-7921
307511	456254	1	Cash loan	F	N	Y	0	171000	370107	20205	319500	Unaccom	Commerc	Secondar	Married	House / a	0.00531	-11961	-4786
307512	456255	0	Cash loan	F	N	N	0	157500	675000	49117.5	675000	Unaccom	Commerc	Higher ed	Married	House / a	0.04622	-16856	-1262
307513																			

FOR AMT_ANNUITY COLUMN



SAVE AGAIN We need to refresh your file with updates. Click Save to try again. Save

Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

https://d.docs.live.net/dd... Version created last time the... 20-05-2023 23:00

Puzzle For Placement.xlsx... Version created last time the... 20-05-2023 12:35

Coder Army Sheet.xlsx [... Version created last time the... 20-05-2023 12:35

akshay_anand_applicatio... Version created last time the... 20-05-2023 22:49

Book1 (version 1).xlsb [A... Version created from the las... 20-05-2023 22:45

https://d.docs.live.net/dd... Version created last time the... 19-05-2023 06:56

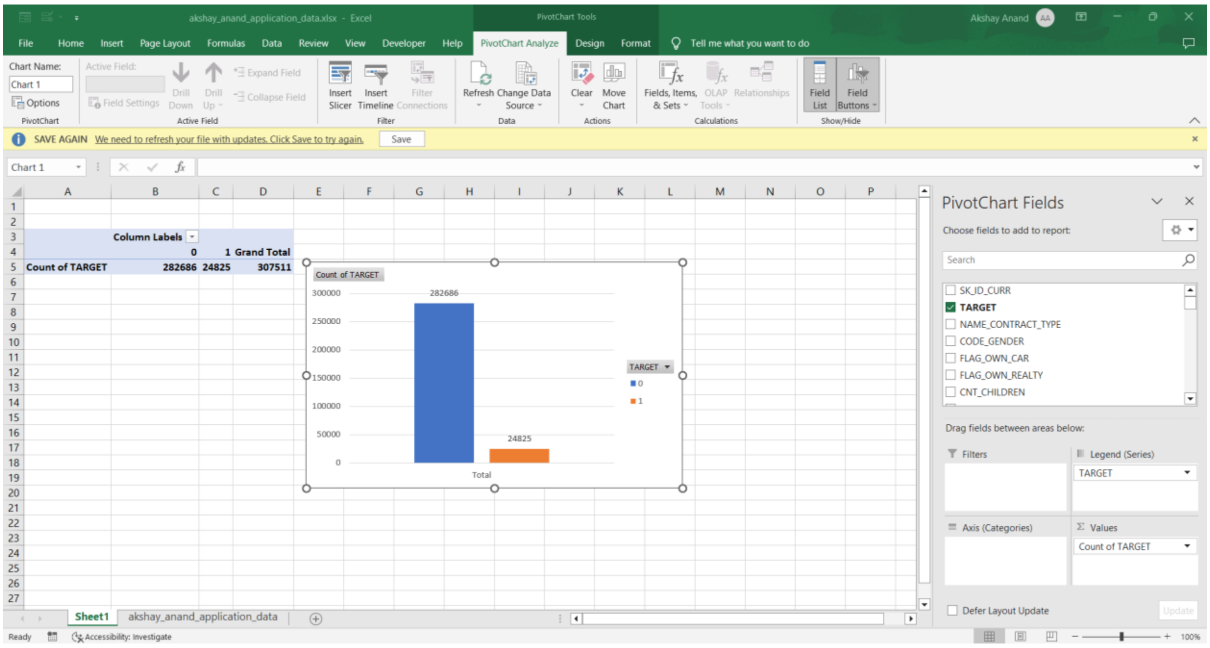
https://d.docs.live.net/dd...

Which file do I want to save?

Close

SO WE SEE THE POINTS WHICH ARE VERY FAR AWAY FROM THE WHISKER PLOT THOSE POINTS ARE OUTLIERS.

- Identify if there is data imbalance in the data. Find the ratio of data imbalance. ***Hint:*** *Since there are a lot of columns, you can run your analysis in loops for the appropriate columns and find the insights.*



akshay_anand_application_data.xlsx - Excel

SAVE AGAIN We need to refresh your file with updates. Click Save to try again.

	Column Labels		
	0	1	Grand Total
Count of TARGET	282686	24825	307511

PivotChart Fields

Choose fields to add to report:

- □ SK_ID_CURR
- ☑ TARGET
- □ NAME_CONTRACT_TYPE
- □ CODE_GENDER
- □ FLAG_OWN_CAR
- □ FLAG_OWN_REALTY
- □ CNT_CHILDREN

Drag fields between areas below:

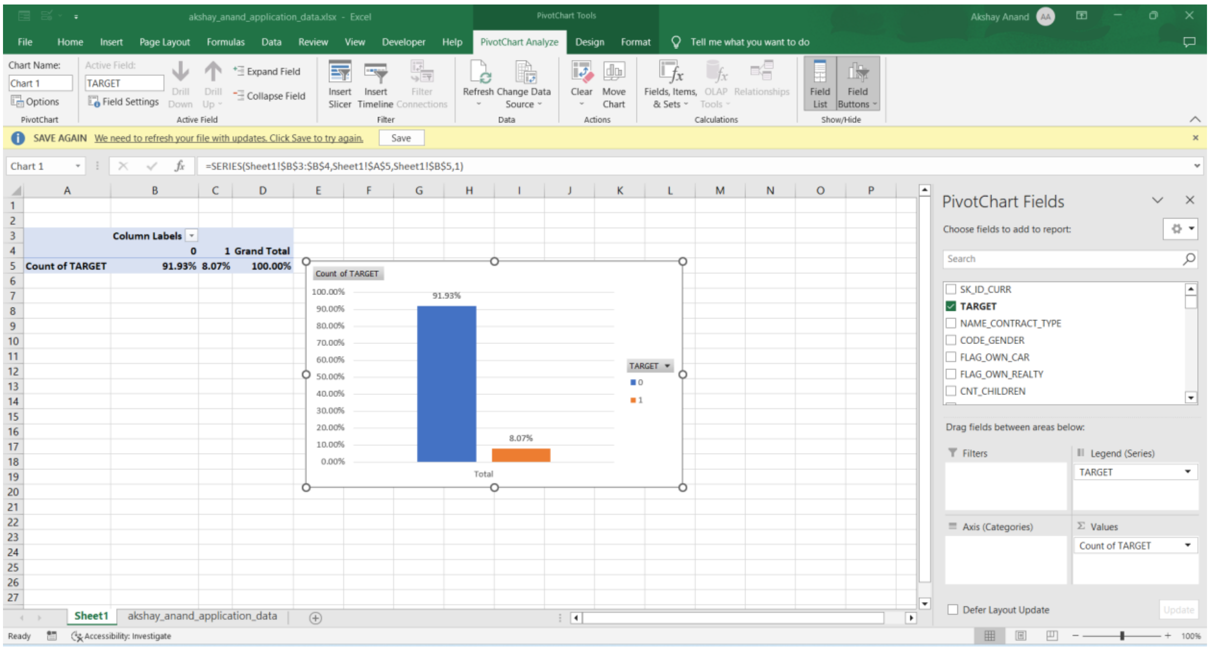
Filters

Legend (Series): TARGET

Axis (Categories)

Values: Count of TARGET

□ Defer Layout Update



akshay_anand_application_data.xlsx - Excel

SAVE AGAIN We need to refresh your file with updates. Click Save to try again.

Active Field: TARGET

=SERIES(Sheet1!$B$3:$B$4,Sheet1!$A$5,Sheet1!$B$5,1)

	Column Labels		
	0	1	Grand Total
Count of TARGET	91.93%	8.07%	100.00%

PivotChart Fields

Choose fields to add to report:

- □ SK_ID_CURR
- ☑ TARGET
- □ NAME_CONTRACT_TYPE
- □ CODE_GENDER
- □ FLAG_OWN_CAR
- □ FLAG_OWN_REALTY
- □ CNT_CHILDREN

Drag fields between areas below:

Filters

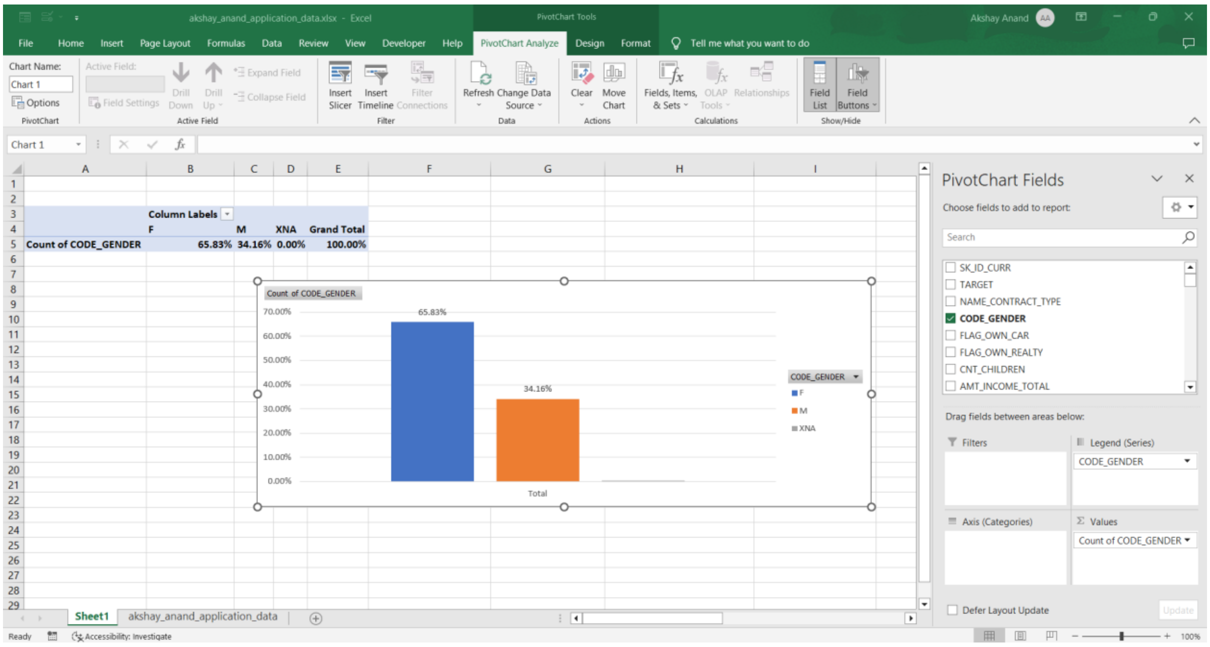
Legend (Series): TARGET

Axis (Categories)

Values: Count of TARGET

□ Defer Layout Update

ALMOST 92 % OF VALUES ARE 0 AND 8% ARE 1. HENCE DATA IMBALANCE IS THERE.



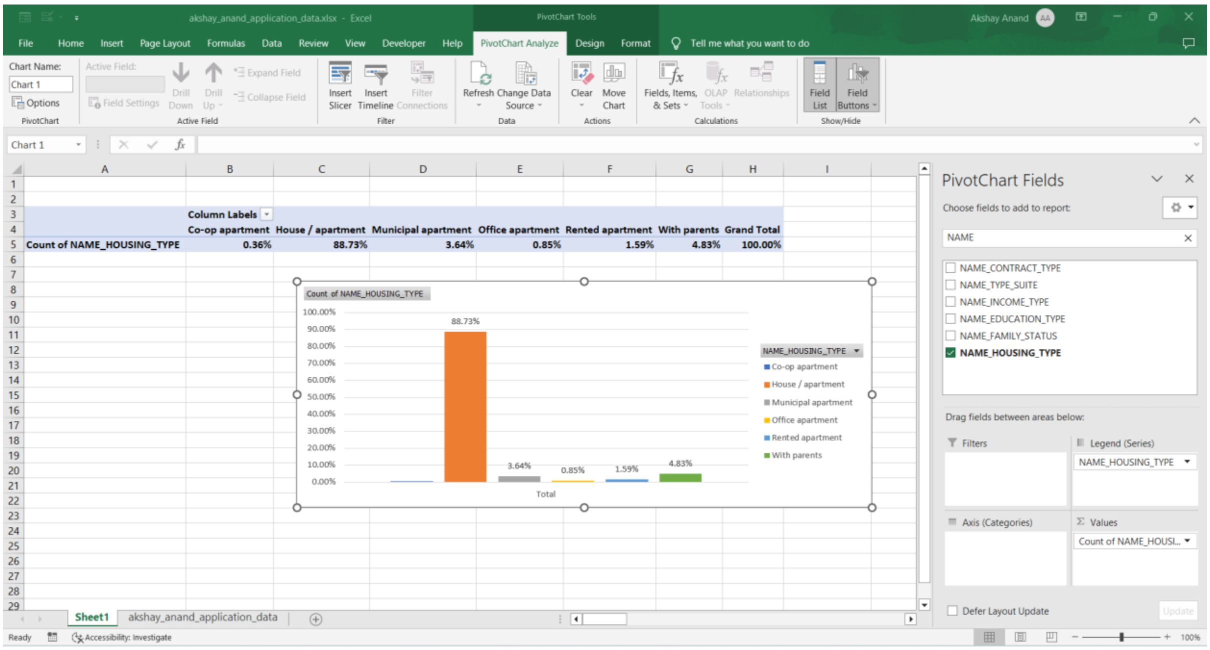
	Column Labels			
	F	M	XNA	Grand Total
Count of CODE_GENDER	65.83%	34.16%	0.00%	100.00%

☑ CODE_GENDER

Legend (Series): CODE_GENDER

Values: Count of CODE_GENDER

FEMALE IS 65%, MALE IS 35%, XNA(BECAUSE OF VERY LESS COUNT) IS 0%. HENCE THERE IS DATA IMBALANCE.



	Column Labels						
	Co-op apartment	House / apartment	Municipal apartment	Office apartment	Rented apartment	With parents	Grand Total
Count of NAME_HOUSING_TYPE	0.36%	88.73%	3.64%	0.85%	1.59%	4.83%	100.00%

☑ NAME_HOUSING_TYPE

Legend (Series): NAME_HOUSING_TYPE

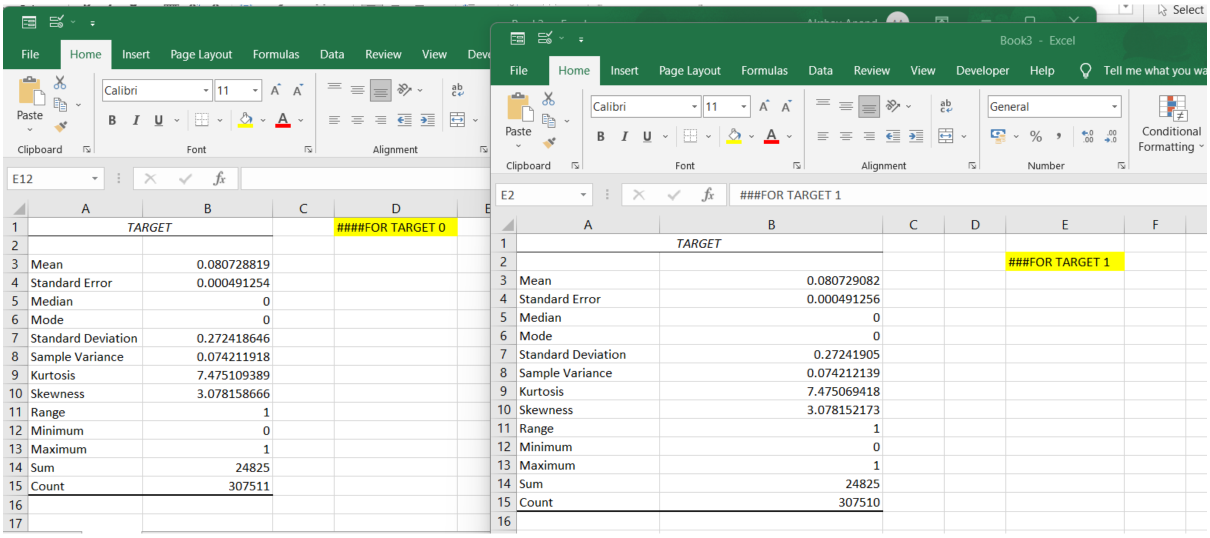
Values: Count of NAME_HOUSI...

HOUSE/APARTMENT'S SHARE IS 89% AS COMPARED TO OTHER HOUSING TYPES. HENCE THERE IS A DATA IMBALANCE.

- Explain the **results of univariate, segmented univariate, bivariate analysis, etc.** in business terms.

WE CAN DO DESCRIPTIVE UNIVARIATE ANALYSIS BY USING THE DATA ANALYSIS TOOL IN THE DATA TAB.

OR WE CAN USE PIVOT TABLES AND GRAPHS FOR UNIVARIATE AND OTHER TYPES OF ANALYSIS.

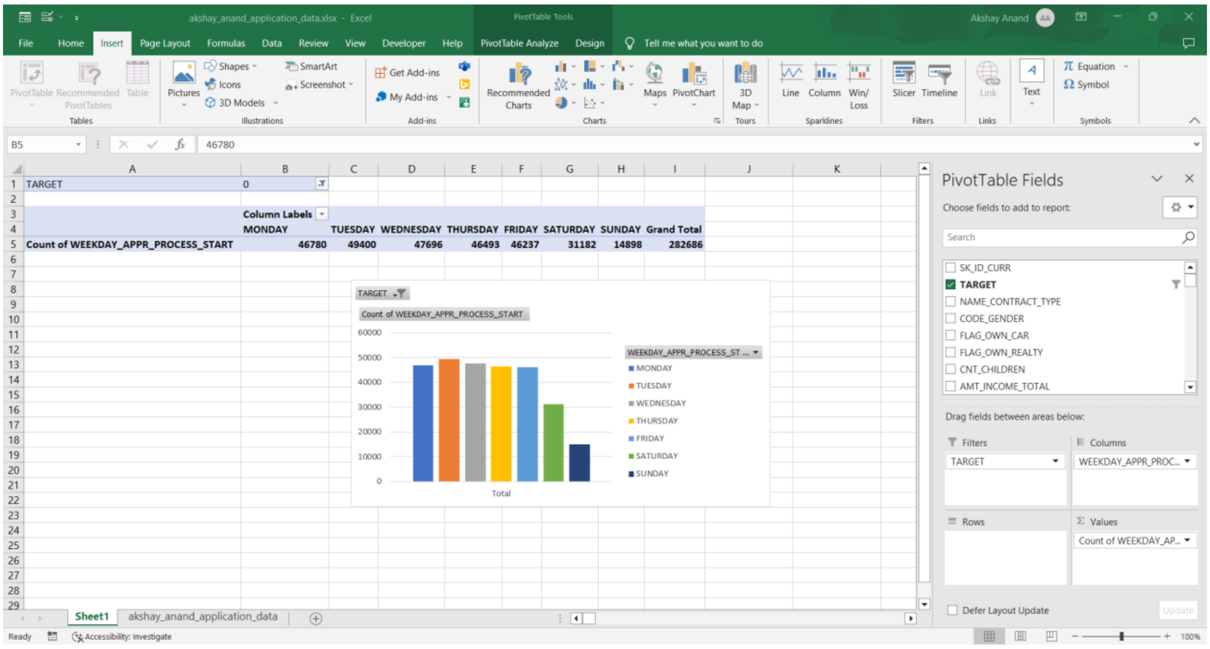


####FOR TARGET 0

TARGET	
Mean	0.080728819
Standard Error	0.000491254
Median	0
Mode	0
Standard Deviation	0.272418646
Sample Variance	0.074211918
Kurtosis	7.475109389
Skewness	3.078158666
Range	1
Minimum	0
Maximum	1
Sum	24825
Count	307511

###FOR TARGET 1

TARGET	
Mean	0.080729082
Standard Error	0.000491256
Median	0
Mode	0
Standard Deviation	0.27241905
Sample Variance	0.074212139
Kurtosis	7.475069418
Skewness	3.078152173
Range	1
Minimum	0
Maximum	1
Sum	24825
Count	307510

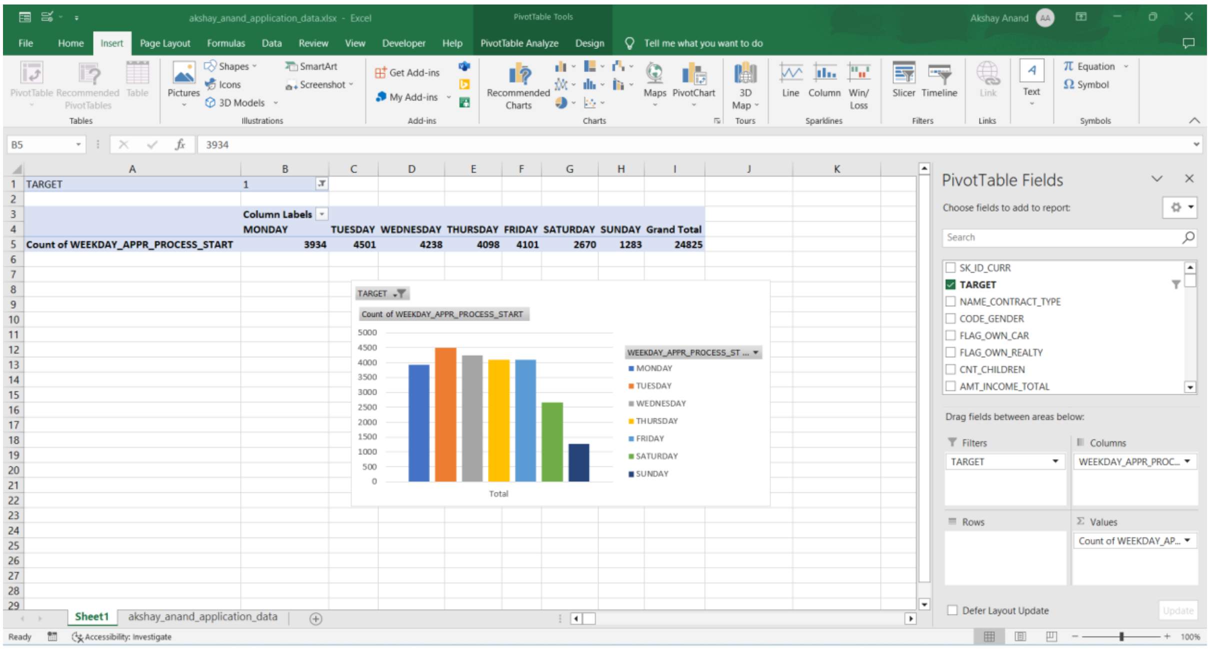


TARGET 0

	Column Labels							
	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY	Grand Total
Count of WEEKDAY_APPR_PROCESS_START	46780	49400	47696	46493	46237	31182	14898	282686

ABOVE IS A UNIVARIATE ANALYSIS FOR WEEKDAY_APPR_PROCESS_START FOR TARGET=0.

SIMILARLY FOR TARGET=1.

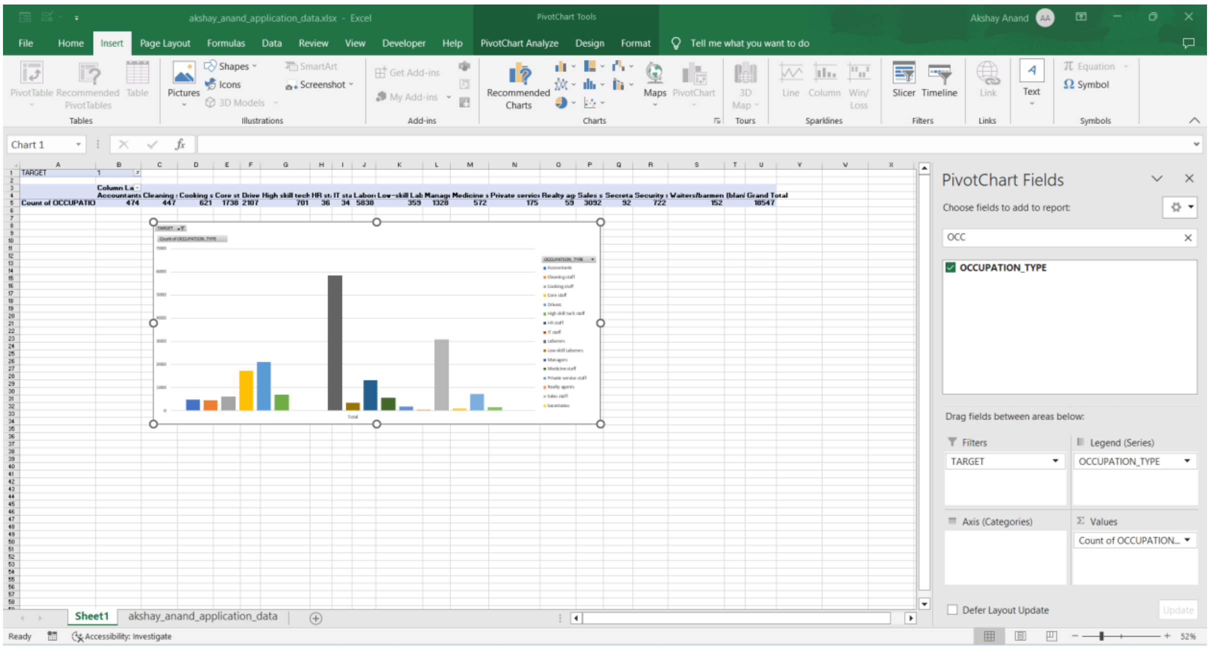


	A	B	C	D	E	F	G	H	I
1	TARGET	1							
2									
3		Column Labels							
4		MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY	Grand Total
5	Count of WEEKDAY_APPR_PROCESS_START	3934	4501	4238	4098	4101	2670	1283	24825

BY THE ABOVE WE CONCLUDE THAT THE APPLICATION STARTING PROCESS IS LESS ON SATURDAY AND SUNDAY IN BOTH CASES OF TARGET = 0 AND 1.

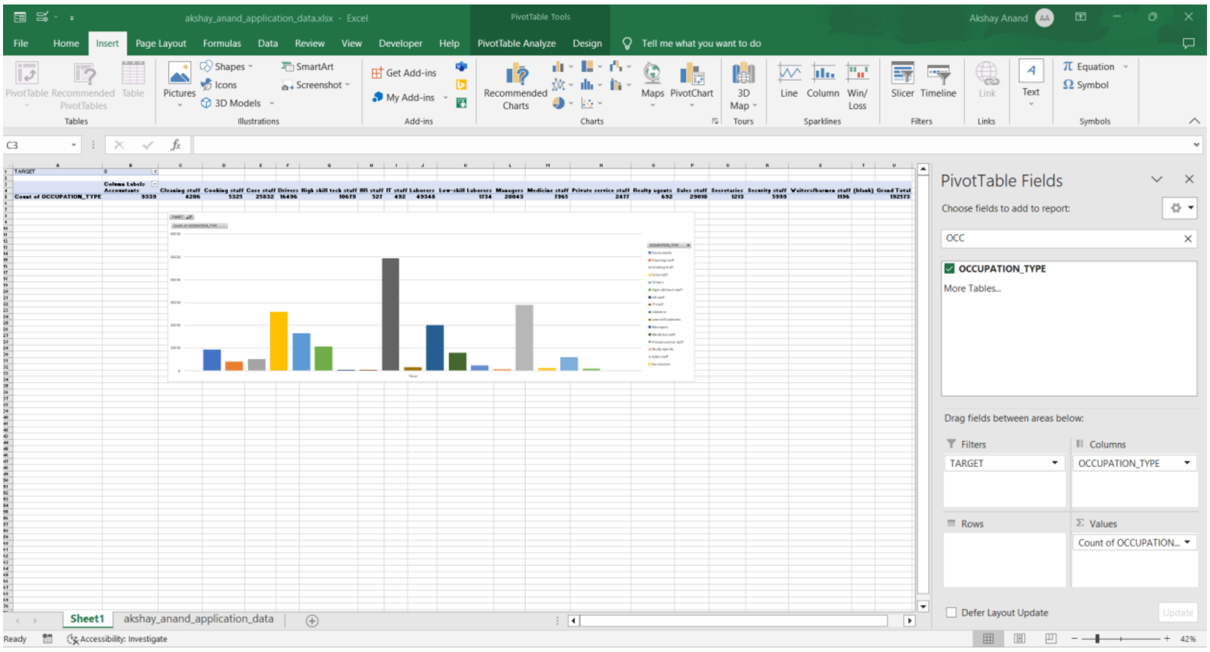
NOW FOR OCCUPATION_TYPE

TARGET=1



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	TARGET	1																			
2																					
3		Column La																			
4		Accountants	Cleaning :	Cooking s	Core st	Drive	High skill tech	HR st	IT sta	Labor	Low-skill Lab	Manag	Medicine s	Private servic	Realty ag	Sales s	Secreta	Security :	Waiters/barmen	(blan	Grand Total
5	Count of OCCUPATIO	474	447	621	1738	2107	701	36	34	5838	359	1328	572	175	59	3092	92	722	152		18547

TARGET=0



TARGET	0

	Column Labels Accountants	Cleaning staff	Cooking staff	Core staff	Drivers	High skill tech staff	HR staff	IT staff	Laborers	Low-skill Laborers	Managers	Medicine staff	Private service staff	Realty agents	Sales staff	Secretaries	Security staff	Waiters/barmen staff	(blank)	Grand Total
Count of OCCUPATION_TYPE	9339	4206	5325	25832	16496	10679	527	492	49348	1734	20043	7965	2477	692	29010	1213	5399	1196		192573

PivotTable Fields

Choose fields to add to report:

OCC

☑ OCCUPATION_TYPE

More Tables...

Drag fields between areas below:

Filters: TARGET

Columns: OCCUPATION_TYPE

Rows:

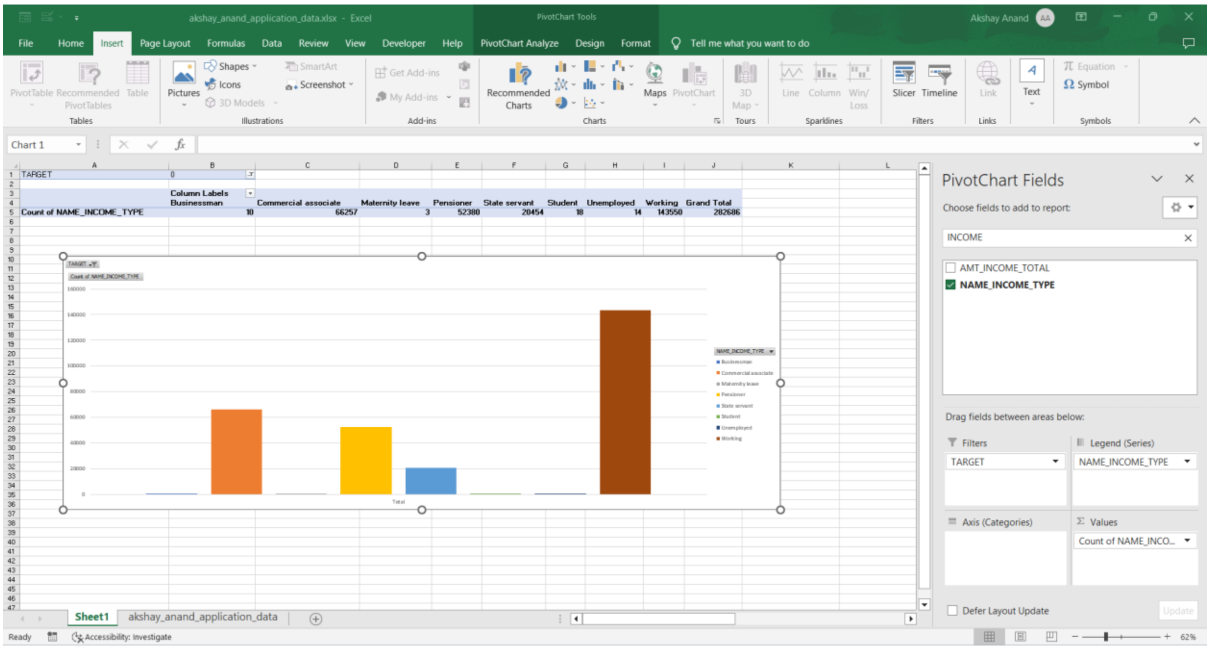
Values: Count of OCCUPATION...

□ Defer Layout Update

IT IS CLEAR FROM THE BAR PLOT THAT CORRESPONDING TO LABOURERS THERE IS A LARGE NO. OF PAYMENT ISSUES IN BOTH CASES OF TARGET= 0 AND 1 WHERE TARGET=0 FOR CLIENTS WITH NO PAYMENT ISSUES AND 1 IS FOR CLIENTS WITH PAYMENT ISSUES.

FOR NAME_INCOME_TYPE

TARGET=0



TARGET	0

	Column Labels Businessman	Commercial associate	Maternity leave	Pensioner	State servant	Student	Unemployed	Working	Grand Total
Count of NAME_INCOME_TYPE	10	66257	3	52380	20454	18	14	143550	282686

PivotChart Fields

Choose fields to add to report:

INCOME

□ AMT_INCOME_TOTAL

☑ NAME_INCOME_TYPE

Drag fields between areas below:

Filters: TARGET

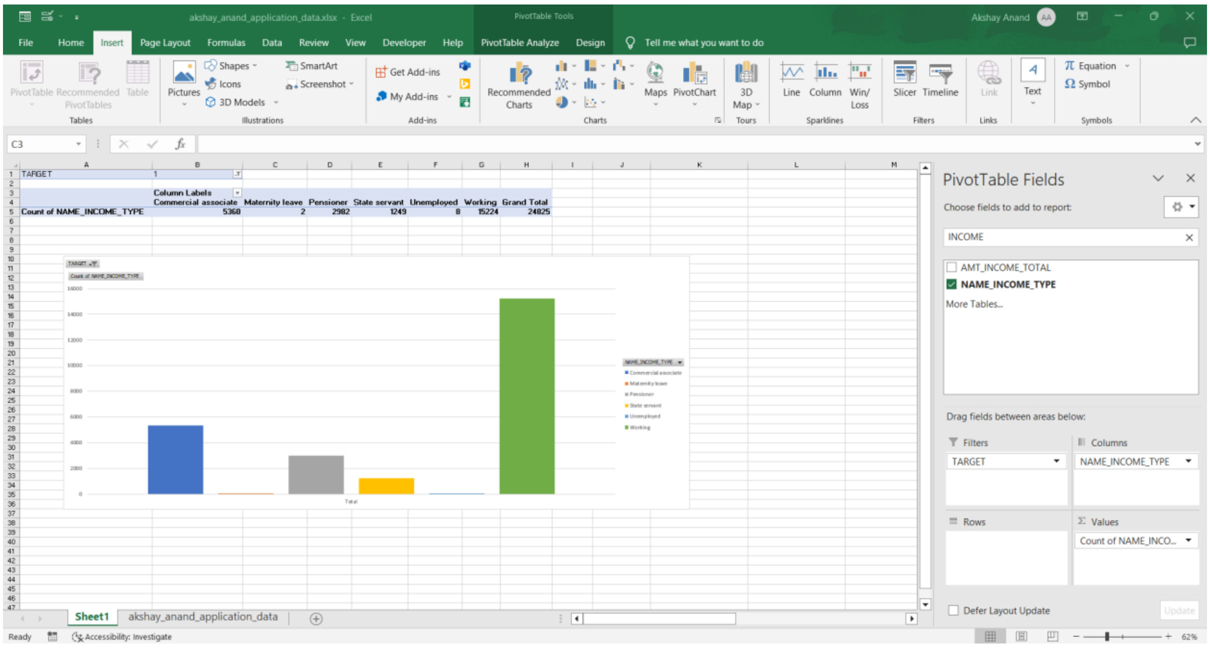
Legend (Series): NAME_INCOME_TYPE

Axis (Categories):

Values: Count of NAME_INCO...

□ Defer Layout Update

TARGET=1



akshay_anand_application_data.xlsx - Excel
PivotTable Tools
File Home Insert Page Layout Formulas Data Review View Developer Help PivotTable Analyze Design
Tell me what you want to do
C3
TARGET
Count of NAME_INCOME_TYPE
16000
14000
12000
10000
8000
6000
4000
2000
0
Total
NAME_INCOME_TYPE
Commercial associate
Maternity leave
Pensioner
State servant
Unemployed
Working
Sheet1 akshay_anand_application_data
Ready
62%

TARGET	1						
	Column Labels						
	Commercial associate	Maternity leave	Pensioner	State servant	Unemployed	Working	Grand Total
Count of NAME_INCOME_TYPE	5360	2	2982	1249	8	15224	24825

PivotTable Fields

Choose fields to add to report:

INCOME

☐ AMT_INCOME_TOTAL

☑ NAME_INCOME_TYPE

More Tables...

Drag fields between areas below:

Filters: TARGET

Columns: NAME_INCOME_TYPE

Rows:

Values: Count of NAME_INCO...

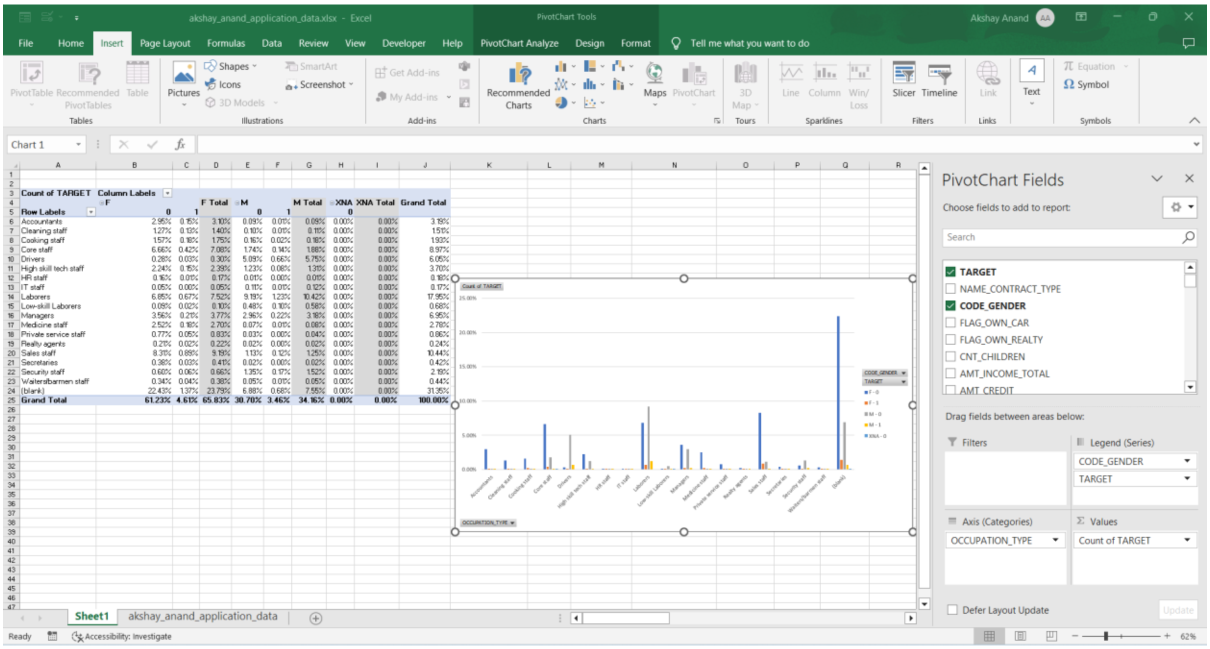
☐ Defer Layout Update

FROM THE ABOVE BAR PLOT WE CAN INFER THAT PEOPLE WITH WORKING TYPE AS WORKING HAVE THE HIGHEST COUNT FOR TARGET =0 AND 1 I.E. FOR CLIENTS WITH AND WITHOUT PAYMENT ISSUES.

SIMILARLY, WE CAN DO UNIVARIATE ANALYSIS FOR OTHER COLUMNS.

UNIVARIATE ANALYSIS IS THE ANALYSIS OF A SINGLE VARIABLE W.R. TO ANOTHER VARIABLE.

SINCE THERE ARE MORE VARIABLES BIVARIATE ANALYSIS IS DONE TO ANALYZE CONCURRENT RELATIONS BETWEEN 2 VARIABLES OR ATTRIBUTES.



akshay_anand_application_data.xlsx - Excel
PivotChart Tools
File Home Insert Page Layout Formulas Data Review View Developer Help PivotChart Analyze Design Format
Tell me what you want to do
Chart 1
Count of TARGET
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
CODE_GENDER
TARGET
OCCUPATION_TYPE
Sheet1 akshay_anand_application_data
Ready
62%

Count of TARGET	Column Labels								
	F		F Total	M		M Total	XNA	XNA Total	Grand Total
Row Labels	0	1		0	1		0		
Accountants	2.95%	0.15%	3.10%	0.09%	0.01%	0.09%	0.00%	0.00%	3.19%
Cleaning staff	1.27%	0.13%	1.40%	0.10%	0.01%	0.11%	0.00%	0.00%	1.51%
Cooking staff	1.57%	0.18%	1.75%	0.16%	0.02%	0.18%	0.00%	0.00%	1.93%
Core staff	6.66%	0.42%	7.08%	1.74%	0.14%	1.88%	0.00%	0.00%	8.97%
Drivers	0.28%	0.03%	0.30%	5.09%	0.66%	5.75%	0.00%	0.00%	6.05%
High skill tech staff	2.24%	0.15%	2.39%	1.23%	0.08%	1.31%	0.00%	0.00%	3.70%
HR staff	0.16%	0.01%	0.17%	0.01%	0.00%	0.01%	0.00%	0.00%	0.18%
IT staff	0.05%	0.00%	0.05%	0.11%	0.01%	0.12%	0.00%	0.00%	0.17%
Laborers	6.85%	0.67%	7.52%	9.19%	1.23%	10.42%	0.00%	0.00%	17.95%
Low-skill Laborers	0.09%	0.02%	0.10%	0.48%	0.10%	0.58%	0.00%	0.00%	0.68%
Managers	3.56%	0.21%	3.77%	2.96%	0.22%	3.18%	0.00%	0.00%	6.95%
Medicine staff	2.52%	0.18%	2.70%	0.07%	0.01%	0.08%	0.00%	0.00%	2.78%
Private service staff	0.77%	0.05%	0.83%	0.03%	0.00%	0.04%	0.00%	0.00%	0.86%
Realty agents	0.21%	0.02%	0.22%	0.02%	0.00%	0.02%	0.00%	0.00%	0.24%
Sales staff	8.31%	0.89%	9.19%	1.13%	0.12%	1.25%	0.00%	0.00%	10.44%
Secretaries	0.38%	0.03%	0.41%	0.02%	0.00%	0.02%	0.00%	0.00%	0.42%
Security staff	0.60%	0.06%	0.66%	1.35%	0.17%	1.52%	0.00%	0.00%	2.19%
Waiters/barmen staff	0.34%	0.04%	0.38%	0.05%	0.01%	0.05%	0.00%	0.00%	0.44%
(blank)	22.43%	1.37%	23.79%	6.88%	0.68%	7.55%	0.00%	0.00%	31.35%
Grand Total	61.23%	4.61%	65.83%	30.70%	3.46%	34.16%	0.00%	0.00%	100.00%

PivotChart Fields

Choose fields to add to report:

Search

☑ TARGET

☐ NAME_CONTRACT_TYPE

☑ CODE_GENDER

☐ FLAG_OWN_CAR

☐ FLAG_OWN_REALTY

☐ CNT_CHILDREN

☐ AMT_INCOME_TOTAL

☐ AMT_CREDIT

Drag fields between areas below:

Filters:

Legend (Series): CODE_GENDER, TARGET

Axis (Categories): OCCUPATION_TYPE

Values: Count of TARGET

☐ Defer Layout Update

ABOVE IS A BIVARIATE ANALYSIS OF OCCUPATION_TYPE VS GENDER W.R. TO TARGET VARIABLE.

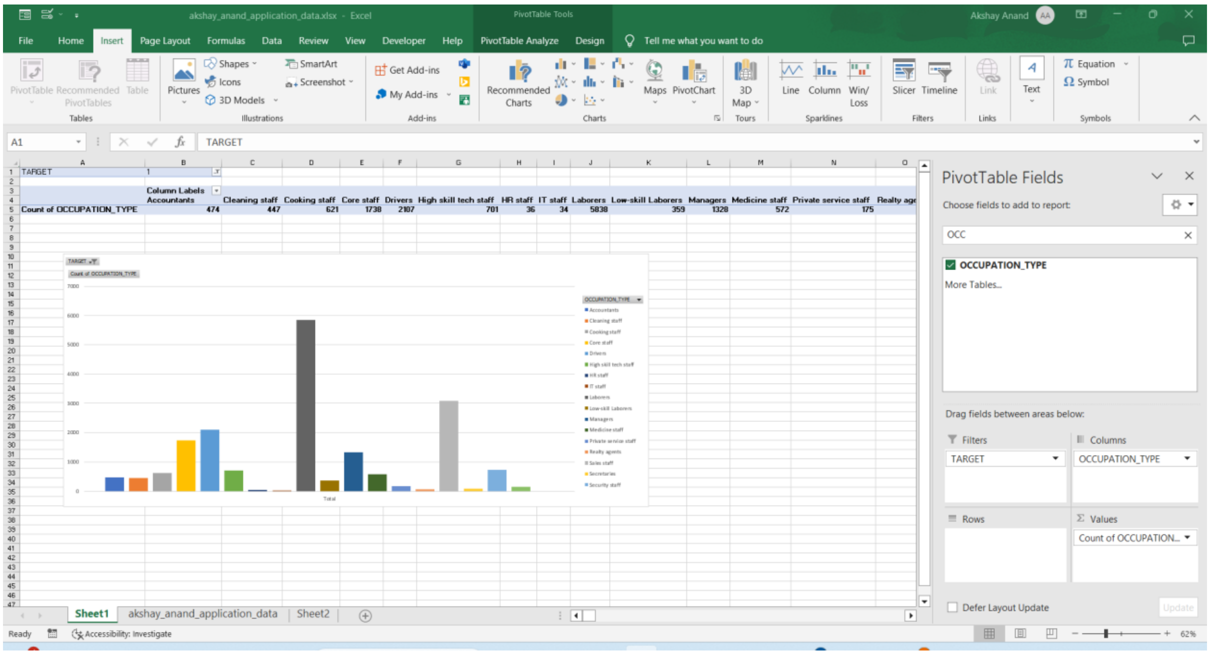
- Find the top 10 **correlation** for the Client with payment difficulties and all other cases (Target variable). Note that you have to find the top correlation by segmenting the data frame w.r.t to the target variable and then find the top correlation for each of the segmented data and find if any insight is there. Say, there are 5+1(target) variables in a dataset: Var1, Var2, Var3, Var4, Var5, Target. And if you have to find top 3 correlation, it can be: Var1 & Var2, Var2 & Var3, Var1 & Var3. Target variable will not feature in this correlation as it is a categorical variable and not a continuous variable which is increasing or decreasing.

FOR TARGET=1 WE FIND A CORRELATION USING THE CORREL FUNCTION BETWEEN AMT_CREDIT AND AMT_CREDIT_TOTAL AND FOUND A CORRELATION COEFFICIENT OF 1.

USED: =CORREL(Sheet1!I2:I307511,Sheet1!H2:H307511)

SIMILARLY WE FINAL CORRELATION COEFFICIENT 1 IN THE CASE OF COUNT_CHILDEREN VS AMT_TOTAL_INCOME. USED: =CORREL(Sheet1!G2:G307511,Sheet1!H2:H307511)

- **Include visualizations** and **summarize** the most important results in the presentation. You are free to choose the graphs which explain the numerical/categorical variables. Insights should explain why the variable is important for differentiating the clients with payment difficulties with all other cases.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	TARGET	1													
2															
3		Column Labels													
4		Accountants	Cleaning staff	Cooking staff	Core staff	Drivers	High skill tech staff	HR staff	IT staff	Laborers	Low-skill Laborers	Managers	Medicine staff	Private service staff	Realty age
5	Count of OCCUPATION_TYPE	474	447	621	1738	2107	701	36	34	5838	359	1328	572	175	

PivotTable Fields

Choose fields to add to report:

OCC

☑ OCCUPATION_TYPE

More Tables...

Drag fields between areas below:

Filters: TARGET

Columns: OCCUPATION_TYPE

Rows:

Values: Count of OCCUPATION_...

☐ Defer Layout Update

Update

LABOURERS' OCCUPATION TYPE HAS THE MOST PAYMENT ISSUES.

OTHER IMPORTANT RESULTS:

1)PEOPLE WITH ACADEMIC DEGREES DO FEWER DEFAULTS.

2)PEOPLE WITH LOW TOTAL INCOME ARE MORE LIKELY TO DEFAULT.

3) PEOPLE WITH HIGH CREDIT AMOUNTS ARE LESS LIKELY TO DEFAULT.

4)THE PROPORTION OF DEFAULTERS I.E. TARGET=1 IS 8% WHILE THAT OF NON-DEFAULTERS I.E. TARGET= 0 IS 92%.

5) PEOPLE BELONGING TO THE WORKING CLASS TEND TO PAY OFF THEIR LOANS ON TIME.

SIMILARLY, BY PERFORMING UNIVARIATE AND BIVARIATE ANALYSES AS DONE ABOVE, WE CAN GET MORE RESULTS.

RESULTS: THROUGH THIS EXPERIMENT, I UNDERSTOOD OUTLIER TREATMENT, UNIVARIATE ANALYSIS, PIVOT TABLES AND PLOTS. I ALSO USED AND UNDERSTOOD THE SIGNIFICANCE OF FILTERING IN EXCEL.