DPDZERO ASSIGNMENT

Sandeep Sethumadhavan

As part of the assignment, I was given an excel file with two sheets: customers and loan details and asked to derive insights from the same. For this data analysis and visualization, I made use of Python. In this document, I go through the specific tasks and highlight the insights I have made. To view the full working code, please refer to the python notebook file which I have attached along with this pdf in the email.

TASKS:

- Use the data provided in data and loan details sheet and join them

The customers sheet contains the data of all the customers and the loan details sheets contains the data of those customers who have taken loans/have loans pending. These two sheets can be joined in different ways.

One would be to just supplement the personal details of the customers who have taken out loans from the customer sheet to the loan details sheet. This would be an inner join and is the straightforward and cleanest option.

The results of this can be seen in the screenshot below. The entire joined dataset can also be viewed by opening the attached excel sheet.

	loan_number	loan_product	amount_pending	due_date	customer_id	firstname	lastname	email	profession
0	100	Unknown	374.69	2020-12-19	440	Felice	Artie	Felice.Artie@yopmail.com	firefighter
1	701	vehicle loan	134.00	2021-05-20	440	Felice	Artie	Felice.Artie@yopmail.com	firefighter
2	824	personal loan	400.62	2021-05-04	440	Felice	Artie	Felice.Artie@yopmail.com	firefighter
3	101	bnpl	688.80	2021-02-14	555	Clementine	Rolf	Clementine.Rolf@yopmail.com	worker
4	825	personal loan	538.56	2021-04-14	555	Clementine	Rolf	Clementine.Rolf@yopmail.com	worker
...	...	...	...	...	...	...	...	...	...
896	811	vehicle loan	499.93	2021-09-09	267	Tierney	Dannye	Tierney.Dannye@yopmail.com	firefighter
897	815	Unknown	100.86	2021-07-09	150	Esmeralda	Velick	Esmeralda.Velick@yopmail.com	police officer
898	816	bnpl	323.68	2021-12-03	101	Alex	Rodmann	Alex.Rodmann@yopmail.com	developer
899	817	bnpl	115.30	2020-09-16	386	Genevra	Wallis	Genevra.Wallis@yopmail.com	worker
900	818	bnpl	126.63	2021-10-18	546	Dianemarie	Elvyn	Dianemarie.Elvyn@yopmail.com	police officer

901 rows × 9 columns

The second way would be to join all the loan details to the customer data. This results in a lot of null values/empty cells for those customers who have not taken out any loans.

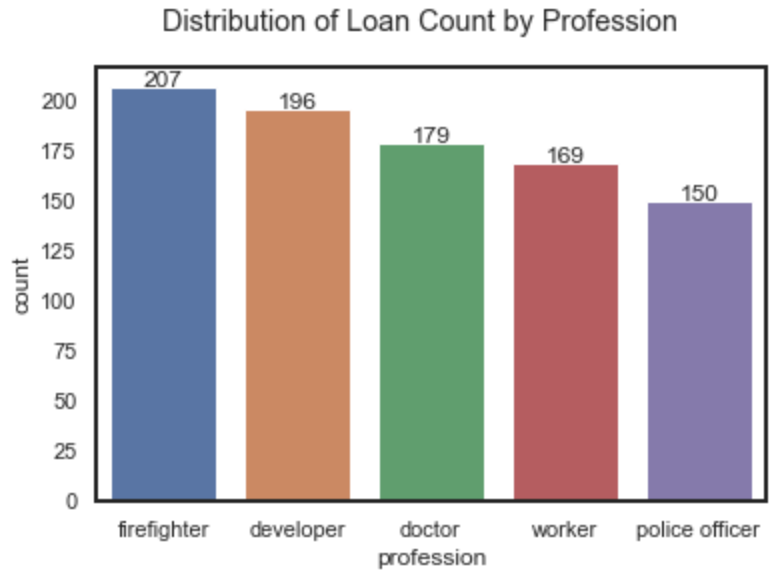
	customer_id	firstname	lastname	email	profession	loan_number	loan_product	amount_pending	due_date
0	100	Tomasina	Marcellus	Tomasina.Marcellus@yopmail.com	police officer	227.0	Unknown	320.20	2020-10-31
1	100	Tomasina	Marcellus	Tomasina.Marcellus@yopmail.com	police officer	316.0	personal loan	329.55	2021-04-14
2	100	Tomasina	Marcellus	Tomasina.Marcellus@yopmail.com	police officer	951.0	personal loan	461.32	2020-08-07
3	101	Alex	Rodmann	Alex.Rodmann@yopmail.com	developer	816.0	bnpl	323.68	2021-12-03
4	102	Diena	Manolo	Diena.Manolo@yopmail.com	firefighter	458.0	personal loan	132.03	2021-02-19
...	...	...	...	...	...	...	...	...	...
1239	819	Rochette	Delacourt	Rochette.Delacourt@yopmail.com	police officer	NaN	NaN	NaN	NaT
1240	820	Anthia	Gale	Anthia.Gale@yopmail.com	doctor	NaN	NaN	NaN	NaT
1241	821	Elvira	Septima	Elvira.Septima@yopmail.com	doctor	NaN	NaN	NaN	NaT
1242	822	Albertina	Quinn	Albertina.Quinn@yopmail.com	police officer	NaN	NaN	NaN	NaT
1243	823	Merrie	Phi	Merrie.Phi@yopmail.com	police officer	NaN	NaN	NaN	NaT

1244 rows × 9 columns

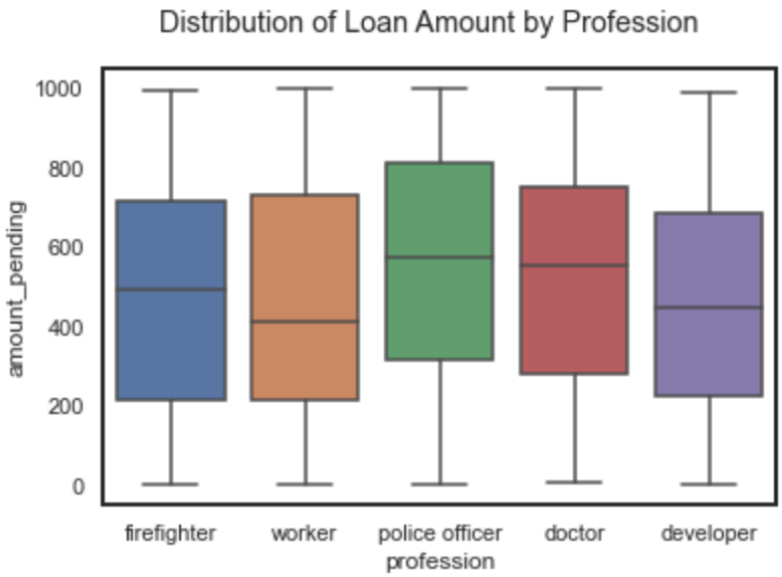
- Create beautiful charts for some of the below items where applicable

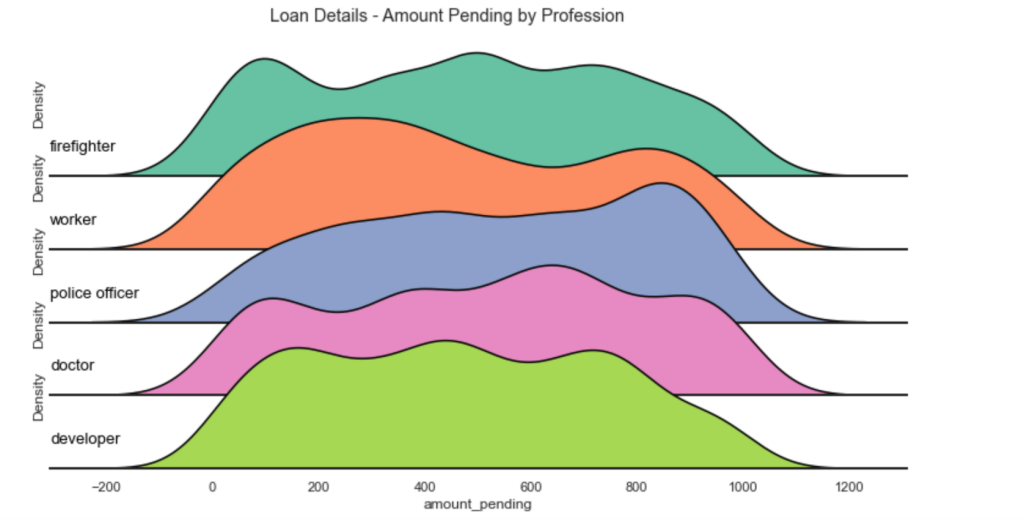
- Distribution of loan count & amount vs profession

First, looking at the distribution of loan count vs the different professions. We can see that firefighters have taken the most number of loans whereas police officers have taken the fewest.



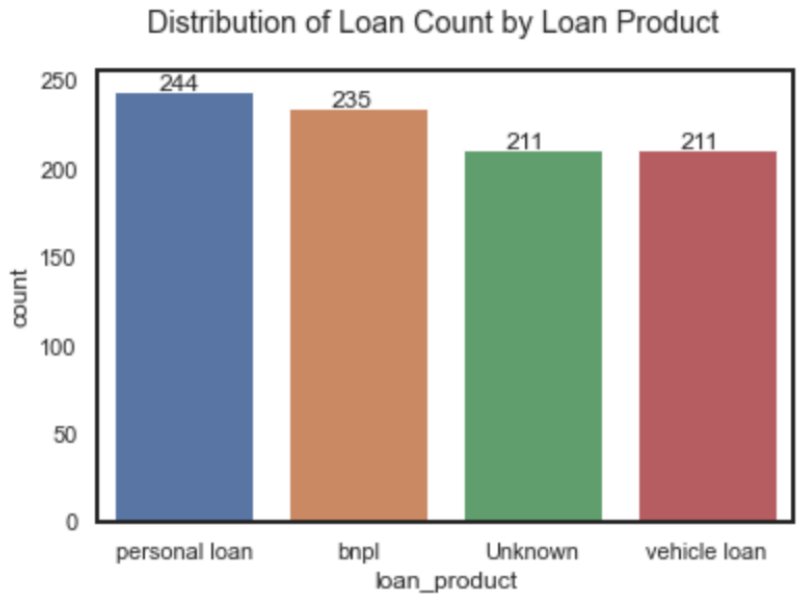
Now studying the distribution of loan amount by profession, it can be inferred that police officers take the biggest loans, followed by doctors. It can also be seen that most workers take smaller loans.



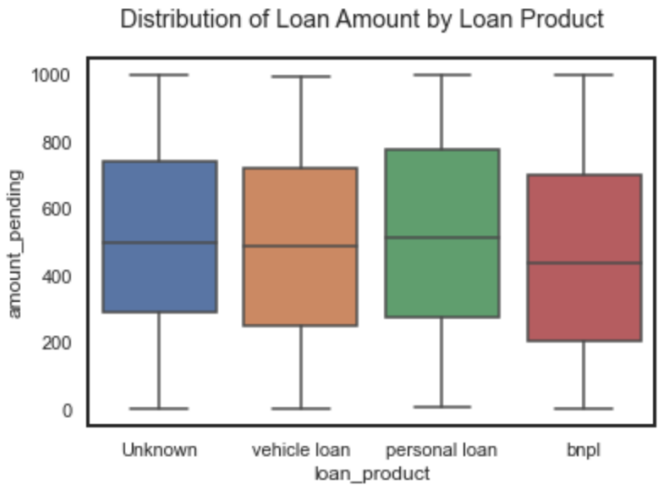


- Distribution of loan count & amount vs loan product

First, looking at the distribution of loan count vs the different loan products. We can see that most of the loans have been personal loans while vehicle loans account for the fewest.



Now studying the distribution of loan amount by loan product, it can be inferred that personal loans account for the biggest loans in terms of the loan amount. It can also be seen that bnpl accounts for the smallest loan amounts.



- *Find out customers who have no loans*

There are 343 customers who have not taken any loans. A sample of this subset can be seen below and the entire dataset can be seen in the attached excel sheet.

	customer_id	firstname	lastname	email	profession
0	100	Tomasina	Marcellus	Tomasina.Marcellus@yopmail.com	police officer
1	101	Alex	Rodmann	Alex.Rodmann@yopmail.com	developer
2	102	Diena	Manolo	Diena.Manolo@yopmail.com	firefighter
3	103	Cristine	Yuille	Cristine.Yuille@yopmail.com	firefighter
4	104	Dania	Auberbach	Dania.Auberbach@yopmail.com	doctor
...	...	...	...	...	...
719	819	Rochette	Delacourt	Rochette.Delacourt@yopmail.com	police officer
720	820	Anthia	Gale	Anthia.Gale@yopmail.com	doctor
721	821	Elvira	Septima	Elvira.Septima@yopmail.com	doctor
722	822	Albertina	Quinn	Albertina.Quinn@yopmail.com	police officer
723	823	Merrie	Phi	Merrie.Phi@yopmail.com	police officer

343 rows × 5 columns

- *Find out customers with max number of loans*

There are a lot of customers who have taken multiple loans. The customers who have taken the ten-most number of loans are displayed below.

	customer_id	firstname	lastname	email	profession	number_of_loans
191	345	Barbara	Pascia	Barbara.Pascia@yopmail.com	developer	9
106	241	Joelly	Wenoa	Joelly.Wenoa@yopmail.com	firefighter	7
267	449	Shaylyn	Hutchison	Shaylyn.Hutchison@yopmail.com	firefighter	7
90	219	Jacenta	Hailee	Jacenta.Hailee@yopmail.com	doctor	7
337	538	Kaja	Anton	Kaja.Anton@yopmail.com	doctor	6
259	438	Kylynn	Urias	Kylynn.Urias@yopmail.com	police officer	6
210	368	Jsandye	Catie	Jsandye.Catie@yopmail.com	doctor	6
5	112	Alejandra	Fontana	Alejandra.Fontana@yopmail.com	doctor	6
276	464	Daphne	Mallon	Daphne.Mallon@yopmail.com	firefighter	6
221	383	Robbi	Gunn	Robbi.Gunn@yopmail.com	police officer	6

- Find out top 10 customer by amount

Taking only individual loans into consideration, given below are the ten customers with the highest loan amount pending.

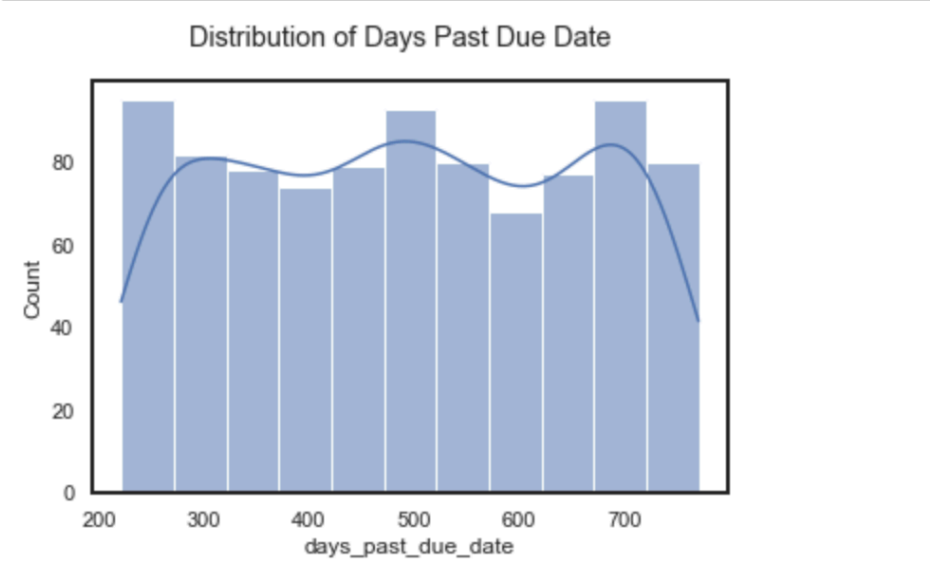
	loan_number	loan_product	amount_pending	due_date	customer_id	firstname	lastname	email	profession
338	927	Unknown	999.79	2020-09-03	114	Raquela	Vharat	Raquela.Vharat@yopmail.com	doctor
55	592	Unknown	999.25	2020-07-26	215	Zondra	Soneson	Zondra.Soneson@yopmail.com	worker
267	903	personal loan	999.10	2021-09-01	156	Lucille	Judye	Lucille.Judye@yopmail.com	police officer
649	359	bnpl	998.63	2021-10-30	127	Isa	Raychel	Isa.Raychel@yopmail.com	worker
171	635	Unknown	998.08	2020-09-30	422	Gusty	Socha	Gusty.Socha@yopmail.com	doctor
656	363	personal loan	996.01	2021-12-27	123	Brooks	Dominy	Brooks.Dominy@yopmail.com	police officer
427	236	bnpl	995.04	2021-03-27	268	Lusa	Merell	Lusa.Merell@yopmail.com	firefighter
627	353	vehicle loan	992.73	2021-03-17	198	Rubie	Bigner	Rubie.Bigner@yopmail.com	firefighter
35	836	bnpl	991.53	2021-08-29	185	Verla	Jehu	Verla.Jehu@yopmail.com	doctor
317	254	bnpl	990.99	2021-02-27	219	Jacenta	Hailee	Jacenta.Hailee@yopmail.com	doctor

As already mentioned above, some of the customers have taken out multiple loans. These loans can be aggregated together to view the top ten customers by total loan amount pending as can be seen below.

	customer_id	firstname	lastname	email	profession	total_amount_pending
106	241	Joelly	Wenoa	Joelly.Wenoa@yopmail.com	firefighter	6231.06
191	345	Barbara	Pascia	Barbara.Pascia@yopmail.com	developer	5407.13
259	438	Kylynn	Urias	Kylynn.Urias@yopmail.com	police officer	4217.75
5	112	Alejandra	Fontana	Alejandra.Fontana@yopmail.com	doctor	3573.94
8	120	Rosaline	Cleo	Rosaline.Cleo@yopmail.com	worker	3315.98
50	174	Mildrid	Alabaster	Mildrid.Alabaster@yopmail.com	firefighter	3236.25
134	277	Philis	Stilwell	Philis.Stilwell@yopmail.com	firefighter	3227.60
6	114	Raquela	Vharat	Raquela.Vharat@yopmail.com	doctor	3203.64
90	219	Jacenta	Hailee	Jacenta.Hailee@yopmail.com	doctor	3142.68
300	492	Rosanne	Irmine	Rosanne.Irmine@yopmail.com	worker	2999.40

- Find out the distribution of DPD (Days past due date) across the loans

By using the due date variable, the days past due date can be calculated. The distribution of this can be seen below. We can see that the distribution is trimodal with peaks at 200, 500 and 700 days.



- Aggregate the data by customer and find out – What's the total amount owed, what is the last due date that has not been paid - what is the avg due date?

The aggregate data for each of the customers including total amount pending, last due date and average due date can be obtained as seen in the table. The entire table can be viewed in the excel sheet attached.

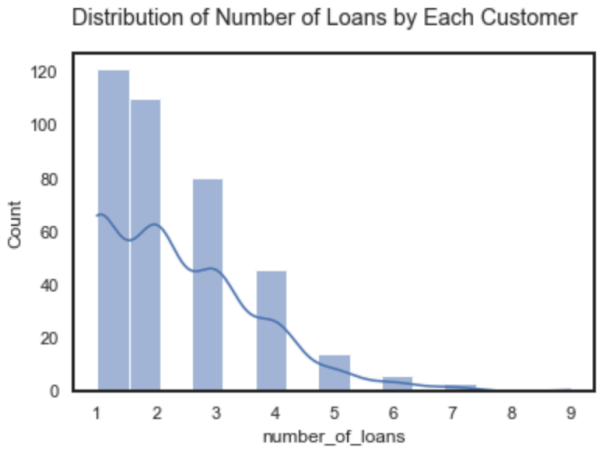
customer_id	firstname	lastname	email	profession	Total_Amount_Pending	Average_Due_Date	Last_Due_Date
100	Tomasina	Marcellus	Tomasina.Marcellus@yopmail.com	police officer	1111.07	2020-11-26 16:00:00	2021-04-14
101	Alex	Rodmann	Alex.Rodmann@yopmail.com	developer	323.68	2021-12-03 00:00:00	2021-12-03
102	Diena	Manolo	Diena.Manolo@yopmail.com	firefighter	673.78	2021-01-08 00:00:00	2021-02-19
103	Cristine	Yuille	Cristine.Yuille@yopmail.com	firefighter	1540.29	2021-09-01 16:00:00	2021-12-21
110	Sashenka	Merell	Sashenka.Merell@yopmail.com	developer	1864.41	2021-04-18 06:00:00	2021-10-09
...	...	...	...	...	...	...	...
593	Brandise	Jerold	Brandise.Jerold@yopmail.com	developer	117.41	2021-01-08 00:00:00	2021-01-08
594	Flo	Natalia	Flo.Natalia@yopmail.com	firefighter	2038.26	2021-02-08 00:00:00	2021-11-07
595	Oralee	Moseley	Oralee.Moseley@yopmail.com	doctor	2657.89	2021-07-21 00:00:00	2021-10-27
597	Hannis	Kravits	Hannis.Kravits@yopmail.com	firefighter	301.27	2020-10-14 12:00:00	2020-10-28
598	Nonnah	Podvin	Nonnah.Podvin@yopmail.com	developer	564.83	2020-08-07 00:00:00	2020-08-07

381 rows × 3 columns

OTHER INSIGHTS

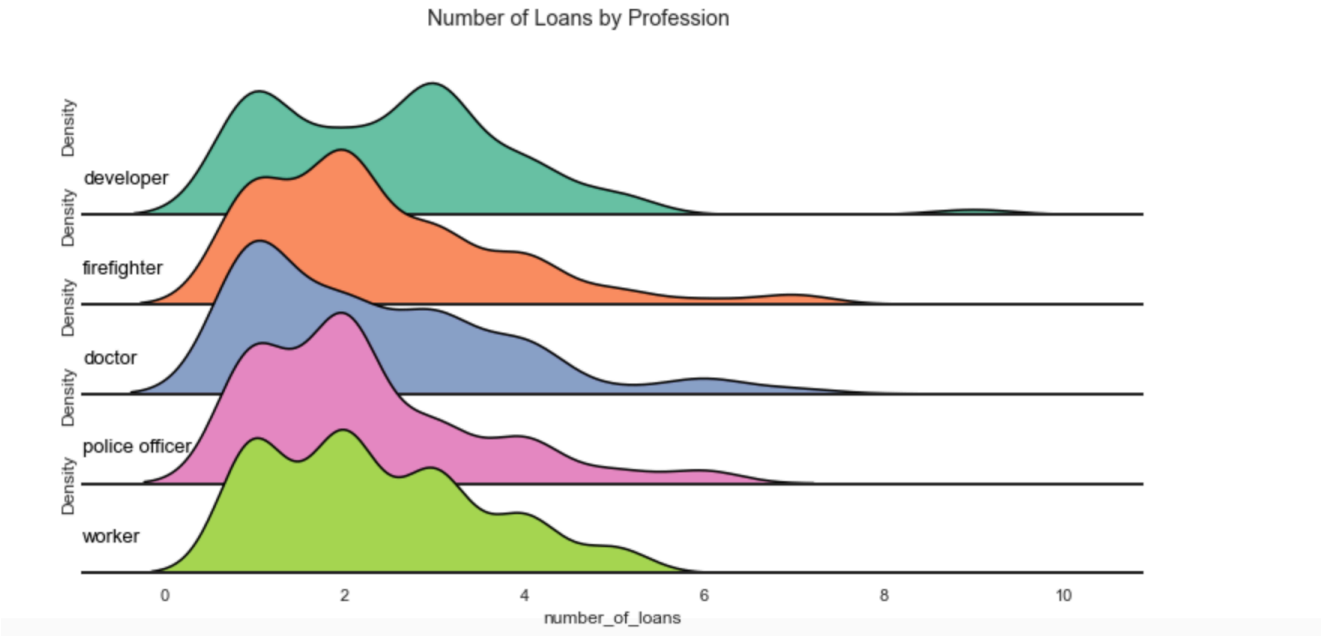
-Distribution of number of loans by each customer:

Most customers seem to have taken 1 or 2 loans and the distribution is heavily skewed to the left.



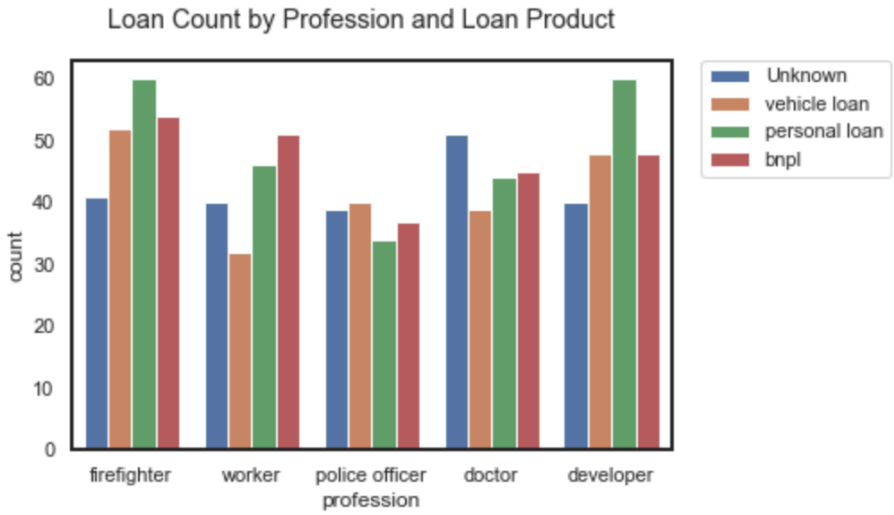
-Distribution of number of loans by profession:

All the professions have left skewed distributions, but it can be seen that individual customers who are developers take out more loans than any other profession.



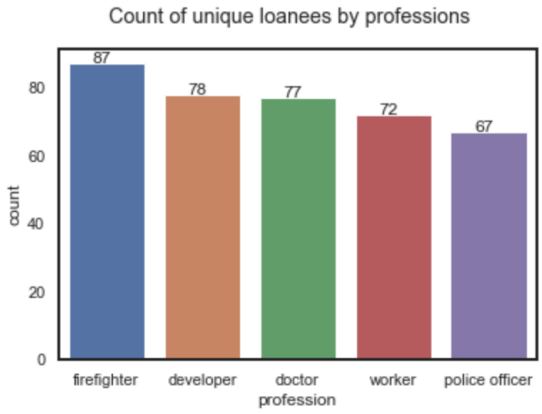
-Distribution of loan count by profession and loan product:

From the below graph, we can see that firefighters and developers mostly take personal loans whereas workers and doctors take bnpl loans mostly. Police officers meanwhile stand out because they take the fewest amount of personal loans and instead mostly take vehicle loans.



-Count of unique loanees by profession:

Here, the customers who have taken out loans have been aggregated and then their counts have been calculated by profession. We can see that the order and even the distribution of the counts are similar to the total loan counts by profession with firefighters taking out the most loans and police officers taking the least.



CONCLUSION

From the data provided, quite a few insights have been made. Out of 724 total customers, only 381 have taken out loans or have loans pending. Of these, firefighters have taken out the most number of loans, followed by developers, doctors, workers; with police officers having taken out the least number of loans. In contrast, police officers owe the biggest loan amounts followed by doctors whereas workers take out the smallest loans.

The loans have been further investigated based on loan product which shows that most loans have been personal loans whereas vehicle loans account for the fewest. In terms of loan amounts, personal loans again account for the biggest loans while bnpl account for the smallest.

343 customers have taken out no loans or have no loan amount pending, whereas a lot of customers have taken out multiple loans with customer number 345 in particular having taken out 9 loans. These multiple loans aggregate to high amounts and customer number 241 has 6231 as the pending amount which is the highest.

Days past due date has also been calculated and it shows that there are a lot of loans which have exceeded considerable number of days past the due date and we have seen that the distribution is trimodal with peaks at 200, 500 and 700 days.

Further information regarding when the loan was issued and interest rates could help us obtain further insights regarding when customers are more likely to take a loan and also how much they will have to pay after interest. Moreover, information regarding loans which have been cleared could also help us see how likely different customers are to pay on time.