

Effect of Inflation on Fraud

IST: 652 Scripting for Data Analysis

Shintaro Osuga

Data:

The data comes from two different kaggle datasets. The first dataset was the Global Inflation Dataset which has 196 different country's inflation information over 44 years and their names. The dataset comes from this kaggle dataset link:

<https://www.kaggle.com/datasets/sazidthe1/global-inflation-data>

The second dataset is the Fraud Detection Transactions Dataset which has 50,000 different transactions in 5 different cities/countries. These transactions have information like what the category which they were used in was, the amount, the account balance, card type, card age, transaction distance, authentication methods, and more.

This data set comes from this kaggle dataset link:

<https://www.kaggle.com/datasets/samayashar/fraud-detection-transactions-dataset>

Data Processing:

For use in this analysis the two datasets must be merged so that the fraudulent activity transactions are linked to their corresponding inflation rate for that country's corresponding year.

The processing is done in the following 5 steps:

1. Changing the city name present in the dataset into a country name as that is what the inflation dataset uses and is what the two will merge on
2. As the fraud dataset contains both fraudulent and non-fraudulent activity, the non-fraudulent activities must be filtered out
3. The fraud dataset contains date time labels as strings, which must be first parsed so that they are datetime objects which can be accessed and utilized later. This is then turned into a new column in the fraud dataset called 'year' which represents the activity year, which will be used later when getting the inflation data for that country for the specific year.
4. The inflation for that country during that year must be retrieved from the inflation dataset. This is done by iterating through the rows of the fraud dataset and utilizing the 'year' column made in the previous step as well as the 'country_name' column made in the first step to grab the correct inflation value from the inflation dataset.
5. The last processing step is to interpret the timestamp string in the fraud dataset into a datetime object which will allow us to interpret it easier and utilize in future analysis.

Once the data is processed with the above steps the 2 datasets are merged into one by merging on the country_name column present in both datasets.

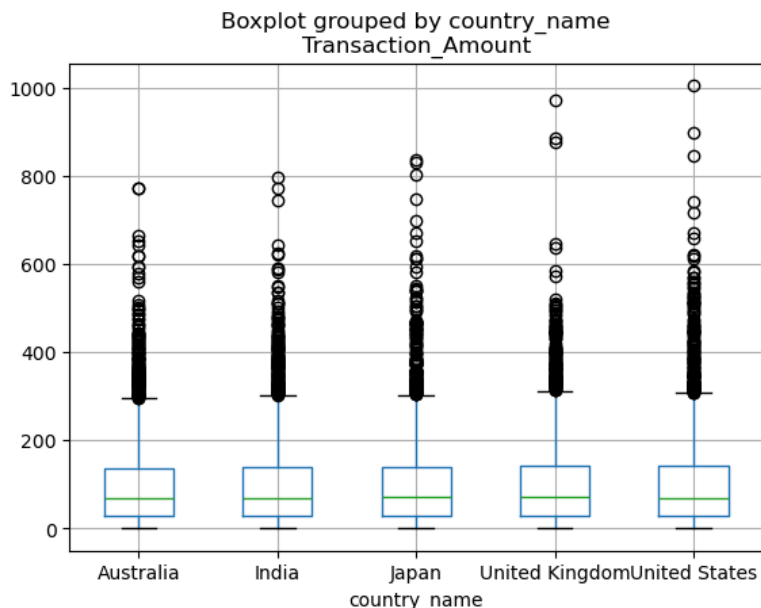
Analysis

Analysis of the differences in transactional amount by country

A good information to know going forth is if there is any difference in the transactional amount between countries. Looking at the table created, the total amount transferred by country is very similar with the largest difference being 17,000, with most of the countries being near or at 310,000-320,000. This is

country_name	
Australia	313112.40
India	311294.35
Japan	330928.09
United Kingdom	321077.68
United States	325205.13

more apparent in the box and whisker plot showing that the range, mean, and standard deviation are about the same for all of the countries but the outliers differ greatly with some countries like the United Kingdom having a very large gap in the outliers. Furthermore, from Australia to the United States in alphabetical order the distance of the furthest outlier increases from just under 800 to just over 1000.



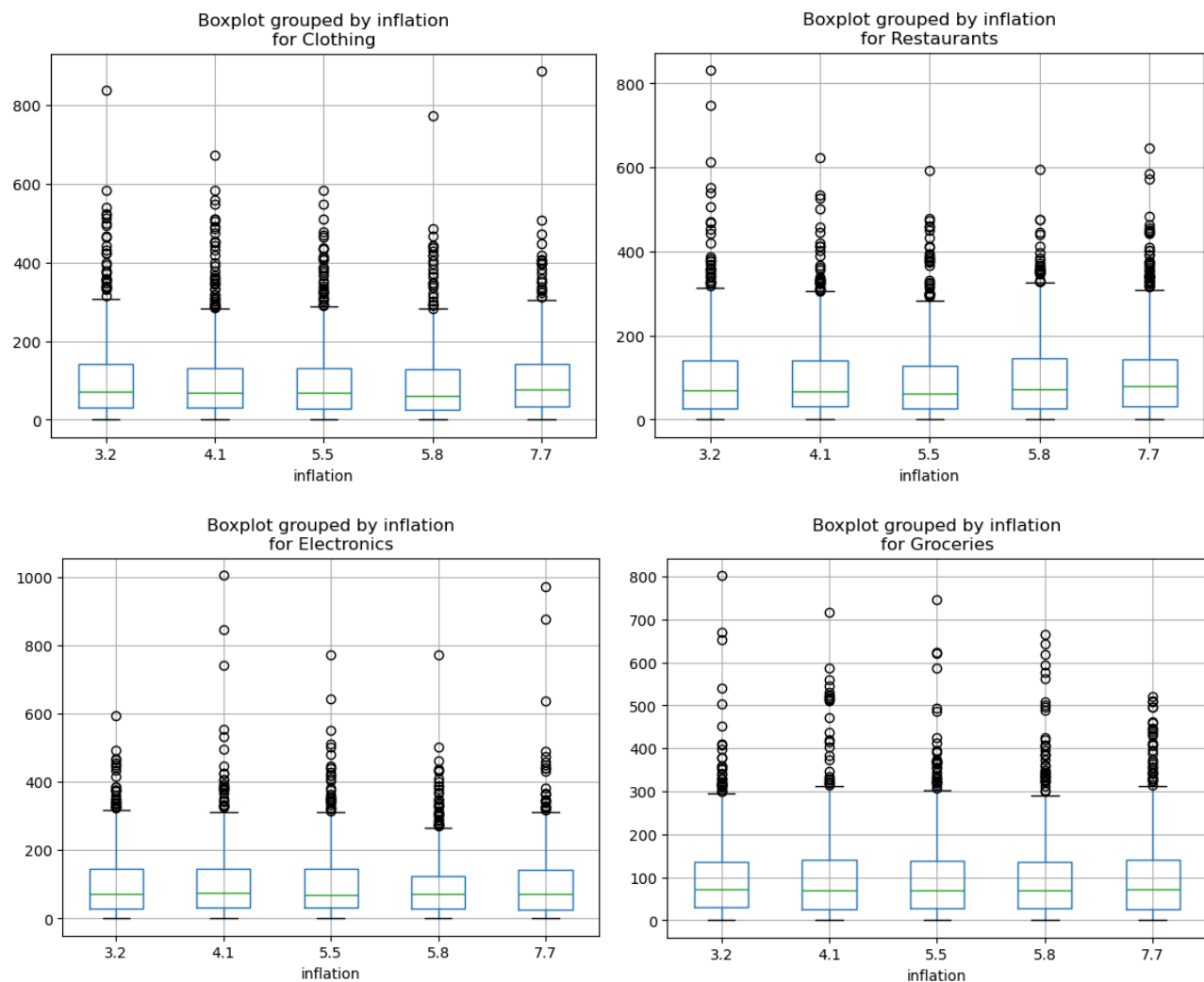
Analysis of the effect of the category of which the fraudulent activity occurred on the rate of inflation

inflation	Merchant_Category	
3.2	Clothing	101.81
	Electronics	100.91
	Groceries	98.68
	Restaurants	102.40
	Travel	95.59
4.1	Clothing	100.99
	Electronics	103.02
	Groceries	99.86
	Restaurants	97.36
	Travel	103.80
5.5	Clothing	97.41
	Electronics	102.90
	Groceries	98.55
	Restaurants	91.76
	Travel	102.24
5.8	Clothing	91.29
	Electronics	94.49
	Groceries	101.69
	Restaurants	100.06
	Travel	102.33
7.7	Clothing	100.09
	Electronics	102.02
	Groceries	101.03
	Restaurants	106.11
	Travel	95.43

The table shows the change in mean transaction amount for each inflation rate and merchant category. As it can be seen that when there is lower inflation there is a lower amount of mean transacted money in the Groceries and Travel categories while during a high inflation there is a higher overall amount of transacted money with only travel having a lower value. Furthermore, when inflation is between 4.1 to 5.8 the lows and highs vary with travel being higher throughout with the others changing throughout.

The 4 plots show each of the different merchant categories and the distribution of the transaction amount by the inflation amount. The small differences in the average outlier distance can be seen to follow the results from the numerical analysis, showing that the overall average transacted amount is around the same but the outliers are

what differentiates the differences between the inflation rates and merchant categories.



Analysis on how inflation affects risk

To analyze the effect of inflation on risk, we must first examine the risk over the different inflation rates. From the produce table it can be seen that the average risk for every inflation rate is around the same, hovering around 0.65 to 0.66. It can be noted though that the inflation rate 5.5 and 5.8 which are very close to one another, also have the 2 highest risk values of 0.67. Knowing that, next a table for the average risk for each inflation and merchant category is made. Shown below, in the table it can be seen that overall every inflation rate as well as the merchant category have a risk score of about 0.66. But it can be noted that when the inflation rate is 5.5 and 5.8 the means for the merchant

inflation		inflation	Merchant_Category	
3.2	0.66	3.2	Clothing	0.66
			Electronics	0.66
			Groceries	0.67
			Restaurants	0.66
			Travel	0.66
4.1	0.65	4.1	Clothing	0.66
			Electronics	0.65
			Groceries	0.65
			Restaurants	0.67
			Travel	0.64
5.5	0.67	5.5	Clothing	0.68
			Electronics	0.68
			Groceries	0.66
			Restaurants	0.67
			Travel	0.67
5.8	0.67	5.8	Clothing	0.65
			Electronics	0.66
			Groceries	0.68
			Restaurants	0.67
			Travel	0.67
7.7	0.66	7.7	Clothing	0.66
			Electronics	0.67
			Groceries	0.66
			Restaurants	0.65
			Travel	0.66

categories slightly increase compared to the other inflation rates. This shows that even though very small there is a slight increase in the risk score for when there is a medium/medium high rate of inflation.

Seeing how there is fluctuation in the risk score over inflation rates the next step is to explore the connections between the risk score and inflation rate by adding more values like card age, transaction distance, account balance, failed transaction count over 7 days, and daily transaction count. Looking at the table below it can be seen that the highest average the risk score reached

⊕ ('inflation', 'Card_Age')	# Risk_Score	# Transaction_Distance	# Account_Balance	# Failed_Transaction_Coun...	# Daily_Transaction_Count
(5.8, 14)	0.91	2035.62	57926.69	2.36	7.21
(5.5, 172)	0.9	1929.42	51239.98	2.83	7.67
(5.8, 174)	0.88	2347.43	72516.57	2.29	6.71
(5.5, 113)	0.88	3339.29	55258.21	3.57	6.43
(3.2, 94)	0.87	2518.4	41861.62	3.12	8.12
(5.8, 77)	0.87	2494.12	43519.91	2.0	6.92
(3.2, 230)	0.87	3293.66	48960.92	1.79	6.79
(3.2, 121)	0.87	1396.02	54390.8	2.6	8.6
(3.2, 112)	0.86	2591.35	58579.17	2.54	7.69
(5.5, 162)	0.86	3032.62	55234.3	2.29	6.47
(4.1, 214)	0.86	2589.44	33528.89	2.5	7.5
(4.1, 143)	0.86	2593.55	55570.61	2.89	6.0
(7.7, 106)	0.86	2219.85	52977.63	2.88	7.62
(3.2, 228)	0.86	2870.64	53247.71	2.64	7.27
(5.5, 202)	0.86	2734.85	41346.94	3.18	7.09
(4.1, 22)	0.85	2932.82	45349.67	2.64	6.0
(5.8, 122)	0.85	2510.94	46620.32	2.62	6.19
(5.8, 27)	0.85	1227.81	55809.9	2.78	4.22
(4.1, 53)	0.85	1961.58	32013.94	2.67	4.83
(5.8, 172)	0.84	2455.83	56373.86	2.53	9.4
(7.7, 2)	0.84	2127.28	46811.08	3.42	9.17
(4.1, 27)	0.84	2809.27	52389.06	2.79	6.0
(5.8, 106)	0.84	2611.35	57426.74	2.94	7.59
(5.5, 175)	0.83	2995.88	53187.7	2.18	6.36
(4.1, 110)	0.83	2701.22	54973.03	2.5	9.42

was 0.91, when sorted by the inflation rate and the card age. This score was reached when the inflation rate was 5.8, and the card age was 14. This activity also has a transaction distance of around 2,000 and an account balance of 58,000 as well as 2.36 failed transactions and 7.21 daily transactions. The parameters of this transaction follow suite with the analysis done above with how the highest average overall risk was attained when the inflation rate was between 5.5 and 5.8.

This shows that inflation and risk is related to inflation and possible other variables as well.

Analysis on the kinds of accounts and activity which are more risky in different inflation rates

Further exploring the types of activities which causes a transaction to be flagged as high or low risk, the variables weekend or not, inflation rate, merchant category, and age of card were used. The risk score was analyzed from these variables, the highest score was attained when the card had an age of 190 used on a weekend on clothing items when the inflation was 4.1. The risk for such a transaction was 0.9999 being just barely under 1 which is the max value it could be. This is interesting as the second highest risk score was attained by an account which on a weekend, bought groceries with a card of age 220, during a period with 3.2% inflation. This transaction

⊞ ('Is_Weekend', 'inflation'...	#	Risk_Score
(1, 4.1, 'Clothing', 190)		0.9999
(1, 3.2, 'Groceries', 220)		0.9999
(0, 5.8, 'Travel', 24)		0.9998
(1, 7.7, 'Groceries', 131)		0.9997
(0, 3.2, 'Travel', 106)		0.9997
(0, 3.2, 'Clothing', 156)		0.9996
(1, 4.1, 'Restaurants', 128)		0.9993
(1, 5.5, 'Travel', 119)		0.9993
(1, 3.2, 'Restaurants', 163)		0.9991
(0, 7.7, 'Groceries', 136)		0.9991
(1, 4.1, 'Restaurants', 129)		0.9991
(1, 7.7, 'Travel', 101)		0.999
(1, 5.5, 'Electronics', 10)		0.999
(0, 5.5, 'Restaurants', 214)		0.999
(0, 3.2, 'Groceries', 151)		0.9989
(1, 4.1, 'Groceries', 185)		0.9989
(1, 5.8, 'Travel', 172)		0.9986
(0, 5.5, 'Electronics', 57)		0.9985
(0, 4.1, 'Clothing', 166)		0.9985
(0, 5.8, 'Electronics', 187)		0.9985
(1, 5.8, 'Clothing', 225)		0.9984
(0, 5.5, 'Clothing', 73)		0.9983
(1, 5.8, 'Travel', 115)		0.9983
(0, 5.5, 'Groceries', 64)		0.9982
(0, 4.1, 'Travel', 40)		0.998

was also 0.9999, but with how commonly people buy groceries over the weekend and the rate of inflation being 3.2 which is not extraordinarily high nor low, shows there must be something more to what makes this such a high risk activity.

Searching deeper, this time the variables chosen were the transaction type, device type, card type, and authentication methods to gain a better understanding of the kind of card and the kind of security which results in a high risk score. A high risk transaction with these variables had an ATM withdrawal transaction type on a tablet with an Amex card having a biometric security when inflation was 5.4. This had a risk score of about 0.75, nearly 0.25 lower than the previous, meaning there must be some activity type which wasn't covered which the top 25% of the highest risk score transactions fall in. Another interesting find

is that the inflation rate for all of the high risk accounts are also during when the inflation was around 5% indicating there might be a correlation between fraudulent activity where accounts with variables like ATM withdrawal, tablet, Amex, biometric have a higher risk score when the inflation rate is 5%.

⊞ ('Transaction_Type', 'Device_Type', 'Card_...	#	inflation	...	#	Risk_Score
('ATM Withdrawal', 'Tablet', 'Amex', 'Biometric')		5.417857142857143			0.7508261904761905
('Online', 'Mobile', 'Amex', 'OTP')		5.342168674698795			0.7417180722891566
('ATM Withdrawal', 'Mobile', 'Mastercard', 'PIN')		5.134615384615385			0.738751282051282
('Online', 'Laptop', 'Discover', 'Password')		5.3820512820512825			0.7347961538461538
('ATM Withdrawal', 'Tablet', 'Amex', 'Password')		5.219753086419753			0.7309135802469136
('POS', 'Mobile', 'Amex', 'Biometric')		5.221875			0.726840625
('POS', 'Tablet', 'Amex', 'Password')		5.234883720930233			0.7265872093023256
('POS', 'Mobile', 'Amex', 'Password')		5.302666666666666			0.7254853333333333
('ATM Withdrawal', 'Tablet', 'Mastercard', 'Password')		5.179310344827586			0.724587356321839
('Bank Transfer', 'Laptop', 'Amex', 'Biometric')		5.054545454545455			0.7242363636363636
('ATM Withdrawal', 'Mobile', 'Visa', 'PIN')		5.117333333333334			0.7210306666666667
('POS', 'Mobile', 'Visa', 'Password')		5.0			0.7188216666666667
('Online', 'Mobile', 'Visa', 'OTP')		5.211702127659574			0.7174021276595745
('Bank Transfer', 'Laptop', 'Discover', 'Password')		4.983333333333333			0.7171694444444445
('Online', 'Tablet', 'Mastercard', 'OTP')		5.173684210526316			0.7171484210526315
('Online', 'Mobile', 'Visa', 'Biometric')		5.4603773584905655			0.7170216981132076
('Online', 'Mobile', 'Amex', 'Biometric')		5.067272727272727			0.7138890909090909
('ATM Withdrawal', 'Tablet', 'Visa', 'OTP')		5.215625			0.7133770833333334
('POS', 'Mobile', 'Amex', 'PIN')		5.381609195402299			0.7133080459770115
('Bank Transfer', 'Tablet', 'Discover', 'Biometric')		5.198823529411764			0.7117482352941177
('POS', 'Laptop', 'Discover', 'PIN')		5.296470588235294			0.7114105882352941
('ATM Withdrawal', 'Tablet', 'Amex', 'PIN')		5.284523809523809			0.7104380952380952
('Bank Transfer', 'Tablet', 'Discover', 'OTP')		5.481944444444444			0.7077305555555555
('POS', 'Tablet', 'Discover', 'Password')		5.258227848101265			0.7075873417721519
('ATM Withdrawal', 'Mobile', 'Mastercard', 'Biometric')		5.351020408163265			0.7054081632653061

⚙️ ('Card_Type', 'inflation', '...' # Risk_Score	
('Discover', 4.1, 3, 5)	0.9977
('Amex', 5.5, 1, 6)	0.9912
('Visa', 3.2, 2, 14)	0.9886
('Visa', 5.8, 3, 14)	0.9803333333333333
('Discover', 5.8, 3, 14)	0.9792666666666667
('Visa', 5.5, 1, 9)	0.979075
('Visa', 7.7, 1, 12)	0.977
('Visa', 7.7, 1, 8)	0.9765
('Visa', 5.8, 2, 10)	0.97635
('Amex', 3.2, 2, 13)	0.9749
('Amex', 7.7, 3, 10)	0.9725
('Discover', 4.1, 3, 8)	0.9722500000000001
('Mastercard', 7.7, 1, 12)	0.9718
('Visa', 5.8, 2, 12)	0.9717
('Mastercard', 4.1, 2, 4)	0.97
('Discover', 5.8, 2, 6)	0.96996
('Amex', 4.1, 0, 2)	0.9699
('Amex', 5.8, 2, 14)	0.9695666666666667
('Visa', 5.8, 2, 9)	0.9691333333333333
('Discover', 3.2, 1, 9)	0.9687666666666667
('Visa', 4.1, 0, 13)	0.96875
('Visa', 3.2, 2, 12)	0.96825
('Discover', 4.1, 0, 6)	0.9676
('Mastercard', 4.1, 0, 6)	0.9672875
('Discover', 5.8, 1, 14)	0.9671125

Combining what we have seen earlier, the variables card type, inflation, failed transaction count over 7 days, and daily transaction count are used. With these variables, the transaction which had the highest risk score was a Discover card with 3 failed transactions in the last 7 days as well as 5 daily transactions when the inflation was 4.1%. This transaction had a risk score of 0.9977 which was marginally higher than the second place which was 0.9912 which had the transaction details of an Amex card with 1 failed transaction in the last 7 days and 6 daily transactions when the inflation was 5.5%. These two show that when the ratio between the number of failed transactions and daily transactions is low then the risk score will be high.

Going further, combining all of the variables utilized earlier, it is found that an activity occurring on a weekday revolving clothing through a bank transfer using a tablet which is on a Discover card that is age 29,

using OTP for authentication and has had 0 previous failed transactions while also have 8 daily transactions when the inflation rate is 5.8% had the highest risk score of 1.0.

⚙️ ('Is_Weekend', 'Merchant_Category', 'Card_Age', 'Transaction_Type', '...' # inflation ... # Risk_Score		
(0, 'Clothing', 29, 'Bank Transfer', 'Tablet', 'OTP', 'Discover', 0, 8)	5.8	1.0
(1, 'Groceries', 220, 'Online', 'Mobile', 'Biometric', 'Mastercard', 3, 4)	3.2	0.9999
(0, 'Restaurants', 170, 'POS', 'Tablet', 'Password', 'Mastercard', 3, 6)	5.5	0.9999
(0, 'Travel', 67, 'POS', 'Laptop', 'PIN', 'Amex', 4, 6)	4.1	0.9999
(0, 'Travel', 12, 'ATM Withdrawal', 'Laptop', 'PIN', 'Amex', 3, 9)	4.1	0.9999
(1, 'Groceries', 9, 'POS', 'Laptop', 'PIN', 'Visa', 2, 7)	7.7	0.9999
(1, 'Clothing', 190, 'POS', 'Laptop', 'Password', 'Discover', 0, 3)	4.1	0.9999
(0, 'Groceries', 82, 'Bank Transfer', 'Tablet', 'PIN', 'Visa', 0, 8)	4.1	0.9998
(0, 'Groceries', 77, 'Online', 'Tablet', 'OTP', 'Discover', 2, 2)	5.8	0.9998
(0, 'Travel', 24, 'ATM Withdrawal', 'Laptop', 'Biometric', 'Amex', 3, 6)	5.8	0.9998

From this analysis of the inflation's effect on risk scores it can be reasonably said that most accounts are vulnerable to attacks and fraudulent activity but most of the higher risk transactions are being done by purchasing clothing when the inflation rate is lower or medium, at about 3-6%.

Results

Overall there were no outstanding results which were found, but with fairly high certainty there was indeed a correlation found between the inflation rate and the risk though the exactness was not discovered it can be reasonably said that when the inflation rate is lower or medium, in the range of 3-6% the risk can be expected to be higher.

As a final thought, the results of this analysis are not to be taken as actual results for the real words as the fraudulent transaction data was synthetically made with noise, which also posed to be a large challenge when exploring for deep and insightful analysis on the data.