



ITESO, Universidad
Jesuita de Guadalajara

Maestría en Ciencia de Datos

Análisis estadístico multivariable

Práctica 1. Covarianza y Correlación

Saúl Enrique Mendoza Ruiz

Mario Alejandro Oviedo Vázquez

30 de enero del 2023

Introducción:

Crecy es una aplicación de préstamos criptográficos basada en las API de Open Banking y Gig Platforms. Ofrece líneas de crédito revolving de hasta \$100/mes que los usuarios pueden acceder vinculando sus cuentas bancarias y/o de trabajo temporal a la aplicación y recientemente ha lanzado un servicio beta de tarjeta de crédito respaldado por criptomonedas.

En este documento realizaremos un análisis estadístico descriptivo en una encuesta realizada a usuarios que poseen tokens del protocolo NEAR.

Objetivo:

Obtener una muestra significativa a través de encuestas que nos permita entender mejor el comportamiento y necesidades de los clientes "NEAR", adaptar los productos, así como dirigirlos de manera más eficiente.

Descripción de datos:

Se realizaron las siguientes preguntas:

- ¿Te gustaría contar con la opción de generar rendimiento en tus activos digitales sin la necesidad de solicitar una línea de crédito con tarjeta Crecy?
- ¿Qué valor de *NEAR en MXN *planeas poner como garantía?
- Prioridad #1
- Prioridad #2
- Prioridad #3
- ¿A cuántos meses te gustaría financiar la mayoría de las compras que hagas con tu tarjeta Crecy?

Para las prioridades se asignaron tres diferentes opciones de respuesta:

- Quiero optimizar mis impuestos.
- Me da oportunidad de acceder a liquidez sin vender mis activos digitales.
- Me gusta que puedo generar rendimientos en mis criptos.

El tamaño de la muestra fue de 126 encuestados.

	Wish yield without CL	NEAR in MXN	P 1	P 2	P 3	Loan lenght		Wish yield without CL	NEAR in MXN	Loan lenght
0	1	16000.0	Taxes	Liquidity	Liquidity	1	count	126.000000	126.000000	126.000000
1	1	1500.0	Liquidity	Yield	Taxes	12	mean	0.880952	13349.841270	6.579365
2	1	5000.0	Yield	Liquidity	Taxes	6	std	0.325137	25911.031249	3.876809
3	1	150000.0	Taxes	Yield	Liquidity	12	min	0.000000	0.000000	1.000000
4	0	500.0	Taxes	Yield	Liquidity	12	25%	1.000000	1000.000000	3.000000
...	50%	1.000000	3000.000000	6.000000
121	1	6000.0	Liquidity	Yield	Taxes	6	75%	1.000000	10000.000000	12.000000
122	1	5000.0	Liquidity	Yield	Taxes	3	max	1.000000	150000.000000	12.000000
123	1	50000.0	Liquidity	Taxes	Yield	3				
124	1	25000.0	Taxes	Yield	Liquidity	1				
125	1	92300.0	Liquidity	Yield	Taxes	6				

126 rows x 6 columns

Análisis estadístico:

Gráfica de pastel:

Análisis de prioridades ponderas por encuestado

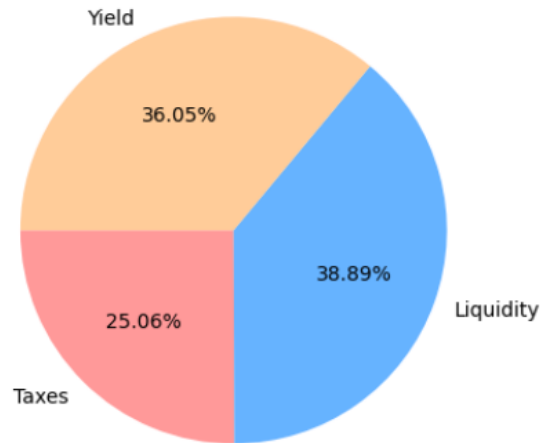
En este ejercicio realizamos un análisis de grafica de pastel para el análisis de las prioridades, para hacer el ejercicio representativo cada voto se dividió entre 7 y se le asigno $\frac{4}{7}$ a la prioridad #1, $\frac{2}{7}$ a la prioridad #2 y $\frac{1}{7}$ a la prioridad #3.

Cabe mencionar que los pesos crecen en una relación de 2 a 1; esto significa que la prioridad 2 tiene el doble de peso que la prioridad 3 y la prioridad 1 tiene el doble de peso que la prioridad 2.

De esta manera podemos conocer las prioridades dando el mismo peso a cada persona.

Para el 38.8% de los encuestados la prioridad fue obtener liquidez, al 36.1% obtener rendimientos y para el 25.1% manejo de impuestos.

Priorities evenly weighted by customers



Análisis de prioridad ponderas por monto.

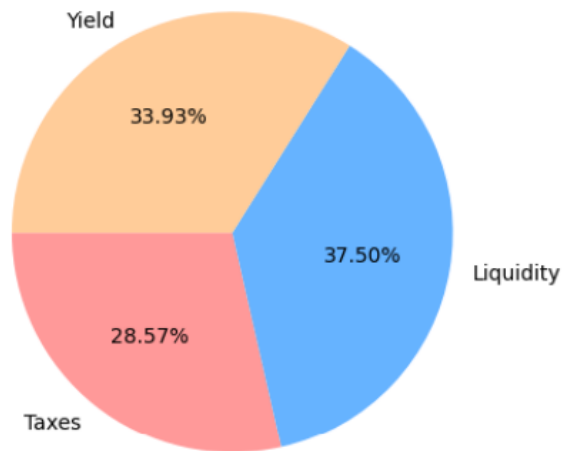
En esta sección dividiremos las cantidades a depositar (MXN) por cada participante por $7; \frac{4}{7}$ partes se asignarán a la primera prioridad, $\frac{2}{7}$ a la segunda y $\frac{1}{7}$ a la tercera.

Como ejemplo un depósito de \$7,000\$ MXN acumulara \$4,000\$ MXN en la etiqueta elegida como prioridad 1, \$2,000\$ en prioridad 2 y \$1,000\$ en prioridad 3.

De esta manera las votaciones de las categorías se ponderan monetariamente.

Podemos destacar que la distribución no presenta variaciones significativas del análisis realizado por prioridad ponderada mostrando que para el 37.5% de los encuestados la prioridad fue obtener liquidez, al 33.9% obtener rendimientos y para el 28.6% manejo de impuestos.

Priorities weighted by deposits



Agrupación de clientes por plazo de crédito

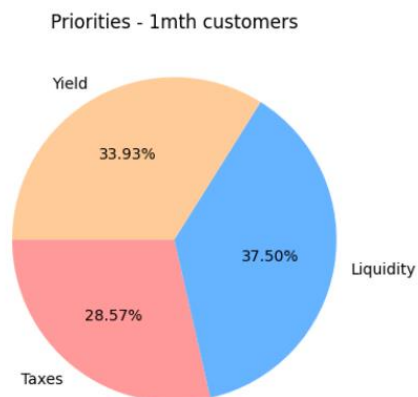
Como primer ejercicio clasificamos las respuestas de las encuestas que desean obtener un rendimiento si la necesidad de solicitar una línea de crédito con base en el plazo de los préstamos que los encuestados pretender solicitar, empezando 1 un mes y siendo el máximo plazo posible 12 meses. La mayoría de los encuestados desean préstamos a 6 meses (34%) seguido de préstamos a 12 meses (27%) y siendo 9 meses el plazo con menos interesados (6%).

	Wish yield without CL													
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%
Loan lenght														
1	16.0	0.812500	0.403113	0.0	1.0	1.0	1.0	1.0	16.0	13906.250000	12318.304469	2000.0	5000.0	10000.0
3	25.0	0.920000	0.276887	0.0	1.0	1.0	1.0	1.0	25.0	11451.600000	23056.131583	5.0	500.0	2000.0
6	43.0	0.930233	0.257770	0.0	1.0	1.0	1.0	1.0	43.0	9564.883721	17672.039575	50.0	1000.0	3000.0
9	8.0	0.875000	0.353553	0.0	1.0	1.0	1.0	1.0	8.0	22635.000000	35141.063892	80.0	3000.0	7500.0
12	34.0	0.823529	0.386953	0.0	1.0	1.0	1.0	1.0	34.0	17085.882353	37030.822798	0.0	625.0	1585.0

Prestamos a 1 mes:

16 encuestados mencionan estar interesados en préstamos con plazo de un mes, cuyo principal motivo es obtener liquidez (37.5%). El monto promedio del préstamo es de \$13,906 MXN, siendo el máximo \$50,000 MXN y el mínimo de \$2,000 MXN con una desviación estándar de \$12

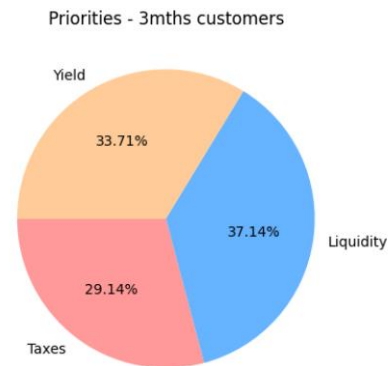
	Wish yield without CL	NEAR in MXN	Loan lenght
count	16.000000	16.000000	16.0
mean	0.812500	13906.250000	1.0
std	0.403113	12318.304469	0.0
min	0.000000	2000.000000	1.0
25%	1.000000	5000.000000	1.0
50%	1.000000	10000.000000	1.0
75%	1.000000	20000.000000	1.0
max	1.000000	50000.000000	1.0



Préstamos a 3 meses:

25 encuestados mencionan estar interesados en préstamos con plazo de tres meses, cuyo principal motivo es obtener liquidez (37.1%). El monto promedio del préstamo es de \$11,451 MXN, siendo el máximo \$100,000 MXN y el mínimo de \$5.00 MXN con una desviación estándar de \$23,056 MXN.

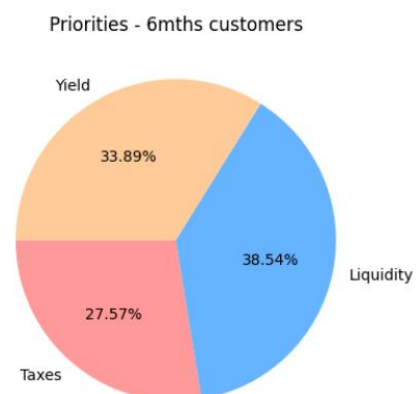
	Wish yield without CL	NEAR in MXN	Loan lenght
count	25.000000	25.000000	25.0
mean	0.920000	11451.600000	3.0
std	0.276887	23056.131583	0.0
min	0.000000	5.000000	3.0
25%	1.000000	500.000000	3.0
50%	1.000000	2000.000000	3.0
75%	1.000000	10000.000000	3.0
max	1.000000	100000.000000	3.0



Préstamos a 6 meses:

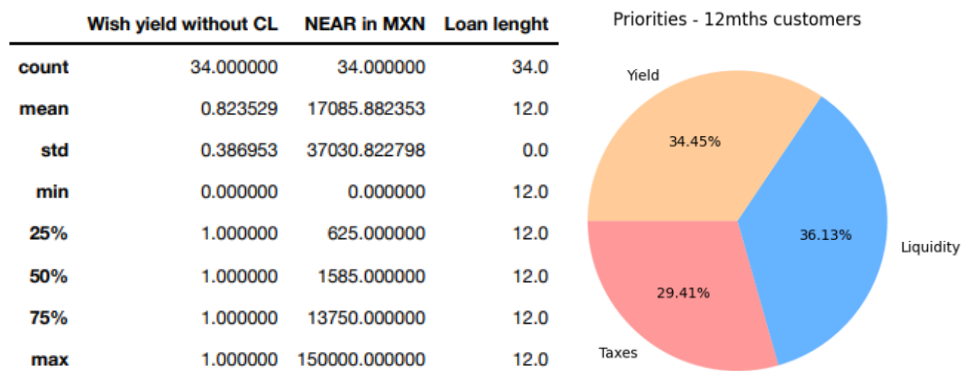
43 encuestados mencionan estar interesados en préstamos con plazo de seis meses, cuyo principal motivo es obtener liquidez (38.5%). El monto promedio del préstamo es de \$9,564 MXN, siendo el máximo \$92,300 MXN y el mínimo de \$50.00 MXN con un desviación estándar de \$17,672 MXN.

	Wish yield without CL	NEAR in MXN	Loan lenght
count	43.000000	43.000000	43.0
mean	0.930233	9564.883721	6.0
std	0.257770	17672.039575	0.0
min	0.000000	50.000000	6.0
25%	1.000000	1000.000000	6.0
50%	1.000000	3000.000000	6.0
75%	1.000000	8000.000000	6.0
max	1.000000	92300.000000	6.0



Préstamos a 12 meses:

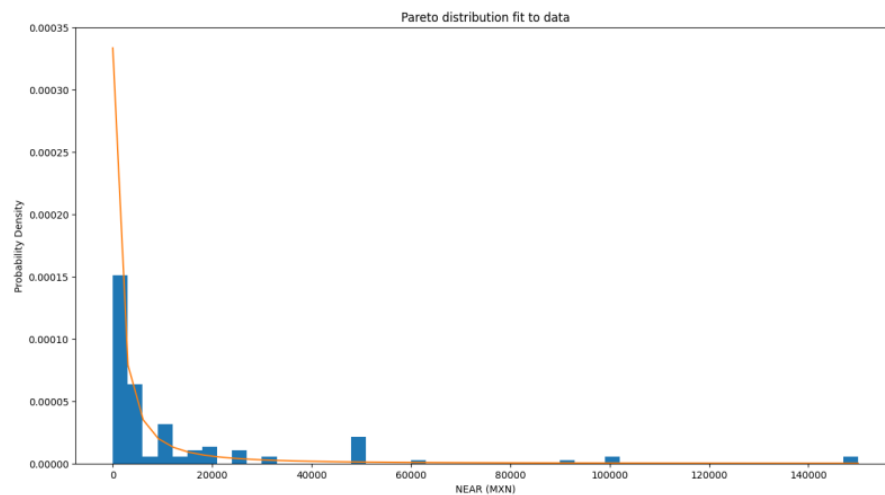
34 encuestados mencionan estar interesados en préstamos con plazo de doce meses, cuyo principal motivo es obtener liquidez (36.1%). El monto promedio del préstamo es de \$17,085 MXN, siendo el máximo \$150,000 MXN y el mínimo de \$0.00 MXN con una desviación estándar de \$37,030 MXN.



Matriz de Correlación:

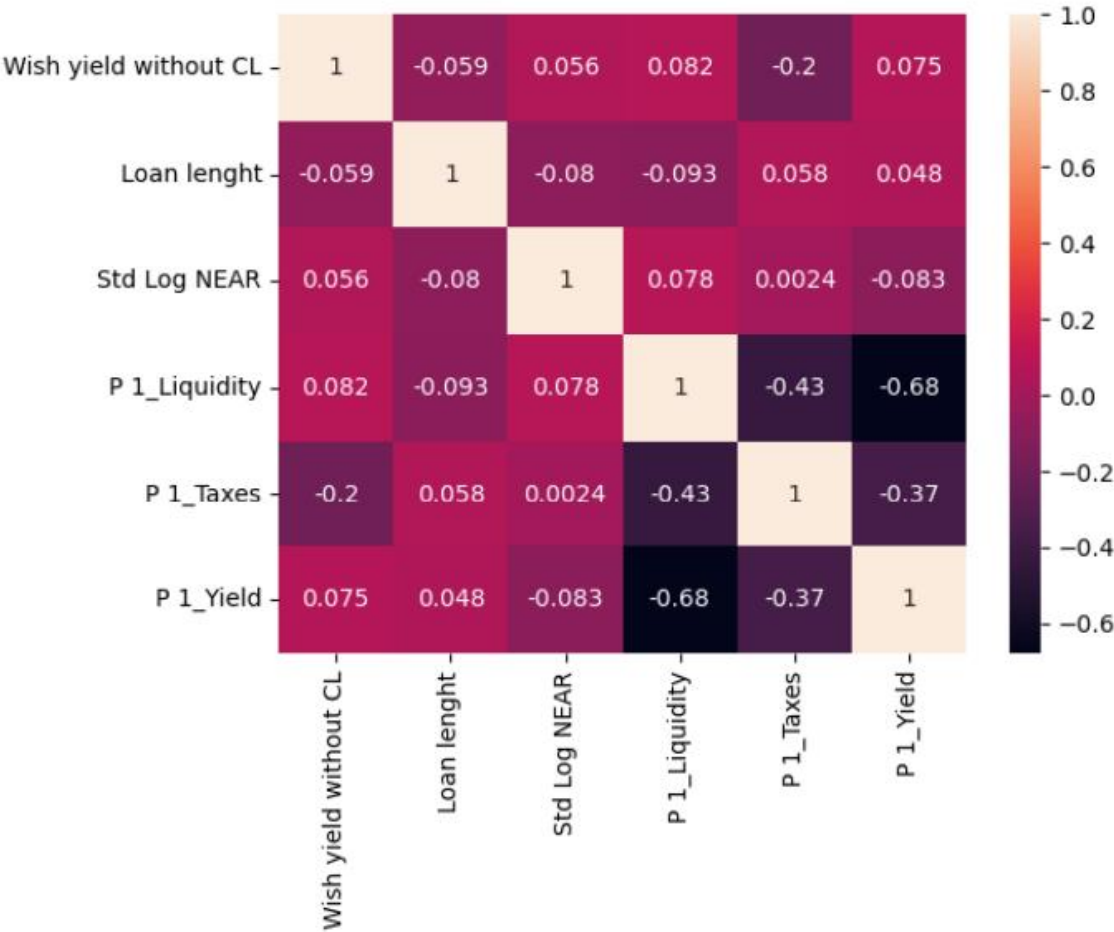
En esta sección escalaremos (función logarítmica) las cantidades de 2NEAR” ya que el comportamiento de las cantidades es “pareto-like”. Luego estandarizaremos los depósitos de “Near” y luego realizaremos un “One Hot Encoding” en las variables “Prioridades”.

Esto debería mejorar la calidad general del análisis de datos.



	Wish yield without CL	Loan lenght	Std Log NEAR	P 1_Liquidity	P 1_Taxes	P 1_Yield
0	1	1	0.812141	0	1	0
1	1	12	-0.299794	1	0	0
2	1	6	0.265724	0	0	1
3	1	12	1.863566	0	1	0
4	0	12	-0.815611	0	1	0
...
121	1	6	0.351371	1	0	0
122	1	3	0.265724	1	0	0
123	1	3	1.347439	1	0	0
124	1	1	1.021802	0	1	0
125	1	6	1.635435	1	0	0

126 rows x 6 columns



Conclusiones:

1. Al agrupar a los clientes a través de diferentes criterios, muestran que su mayor apetito es por la liquidez.
2. Los tenedores no están tan preocupados por los 'Impuestos', existe una correlación lineal negativa de 0.2 entre el deseo de acumular (staking) rendimiento contra el deseo de optimizar los impuestos.
3. La liquidez y el rendimiento están fuertemente correlacionados negativamente (-0,68).
4. No se puede decir que los clientes estén segmentados por poder adquisitivo, las correlaciones son cercanas a cero. Después de escalar y estandarizar, la correlación sigue siendo minúscula.
5. Los encuestados con tokens NEAR muestran una distribución de cola pesada (y un sesgo positivo) con respecto a los depósitos.

Repo: https://github.com/MarioOvz/DS_AEM