



# UNIVERSIDAD DE GRANADA

TRATAMIENTO INTELIGENTE DE DATOS

## Práctica 1

---

**Autor**

Antonio José Muriel Gálvez



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE TELECOMUNICACIÓN

Granada, 16 de octubre de 2024

# Índice

<b>Ejercicio 1</b>	<b>3</b>
a) Algunas variables son categóricas, sin embargo el lector de ficheros Excel de KNIME los reconoce como de tipo entero. Esto ocurre con la variable Family y Education. Utiliza los nodos de cambio de tipo de dato en la carpeta Manipulation/Column/Convert & Replace para que cada variable esté correctamente tipificada. . . . .	4
b) Utiliza el nodo Data Explorer para una visualización y estudio básico de las variables de la base de datos. ¿Existe alguna característica que te llame la atención? ¿Por qué? . . . .	5
c) Realiza un diagrama de burbujas para analizar el número de miembros de la familia de las personas recogidas en la tabla. Dibuja un diagrama de barras de esta característica.	6
d) Realiza un diagrama de dispersión para analizar visualmente los campos edad y años de experiencia. ¿Que opinas sobre la gráfica? Estudia el coeficiente de correlación entre dichas variables. . . . .	8
e) Estudia las matrices de dispersión del resto variables. Calcula la correlación lineal para las variables en las que eso tenga sentido. ¿Sería razonable eliminar alguna característica en base a este estudio? . . . . .	8
f) Asigna colores a las filas en al número de componentes de la familia. A partir de esto, crea un diagrama de dispersión entre la edad y el salario donde los puntos tengan un color dependiendo del número de componentes de la familia. ¿Se puede observar algo en este diagrama? . . . . .	11
g) Filtra la base de datos según la variable Family (nodo Row Filter). Por ejemplo, considera los clientes con familias de 4 miembros (o cualquier otro número, si lo prefieres). Realiza un estudio parecido a los puntos anteriores. ¿Se puede observar algo destacable para estos clientes? . . . . .	12
h) De la base de datos filtrada anterior, calcula diagramas circulares para las variables nominales. . . . .	12
<b>Ejercicio 2</b>	<b>17</b>
a) El nombre de las variables es demasiado largo. Cámbialos por otros más cortos utilizando el nodo Column Rename. . . . .	19
b) Calcula las medidas básicas utilizando el nodo Statistics. ¿Existe algún dato que te llame la atención? . . . . .	20
c) Realiza diagramas de dispersión de las variables numéricas de tipo real (double). Al observar las gráficas, ¿consideras que algunas variables están relacionadas? ¿Cuáles? ¿Existe alguna explicación razonable para esto? . . . . .	22
d) Calcula matriz de correlación de las variables numéricas utilizando el nodo Linear correlation. Compara con el punto anterior. . . . .	23
e) Realiza un diagrama de barras de la columna POS (posición del jugador). . . . .	24
f) Realiza diagramas de cajas de las variables numéricas. ¿Existen jugadores que consideras que sobresalen del resto en alguna característica (outliers)? . . . . .	25

- g) Calcula un diagrama circular de la columna GP (partidos jugados). . . . . 26
- h) Selecciona un equipo y filtra la tabla para obtener los jugadores que pertenecen a dicho equipo, usando el nodo Row Filter. Selecciona 5 características numéricas y filtra la tabla para quedarse únicamente con esas columnas (además de la columna con el nombre). Con la tabla resultante dibuja un diagrama de coordenadas paralelas (dibuja la líneas paralelas en colores a partir de la columna del nombre). ¿Es posible observar algún jugador con unas características diferentes del resto? Prueba con otros equipos y otras características, si no ha sido posible. . . . . 29
- i) Si tuvieras que dividir los jugadores en varios grupos, basándote en los estudios realizados, ¿cómo lo harías? ¿En cuántos grupos? ¿Cómo denominarías/describirías a cada grupo? . 33

# Ejercicio 1

(6 puntos) Considera los datos de la tabla prestamo.xls.

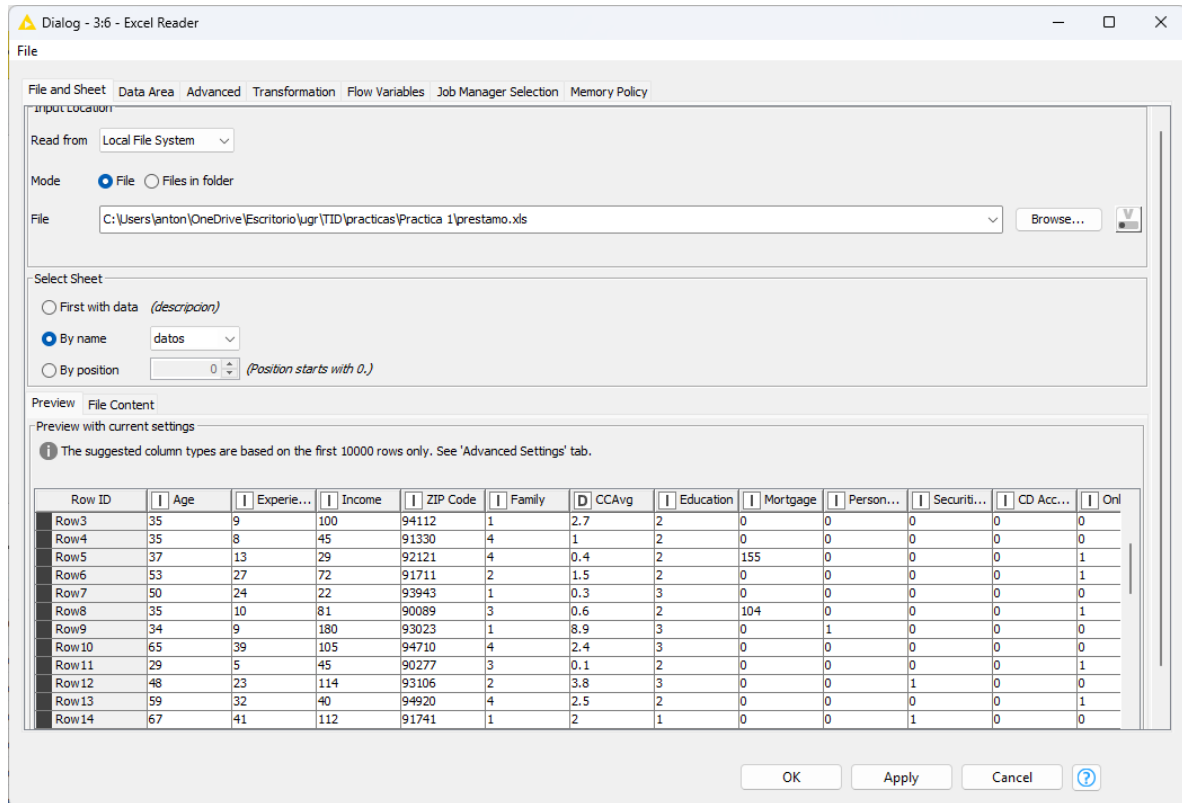


Figura 1: Excel Rader

a) Algunas variables son categóricas, sin embargo el lector de ficheros Excel de KNIME los reconoce como de tipo entero. Esto ocurre con la variable Family y Education. Utiliza los nodos de cambio de tipo de dato en la carpeta Manipulation/Column/Convert & Replace para que cada variable esté correctamente tipificada.

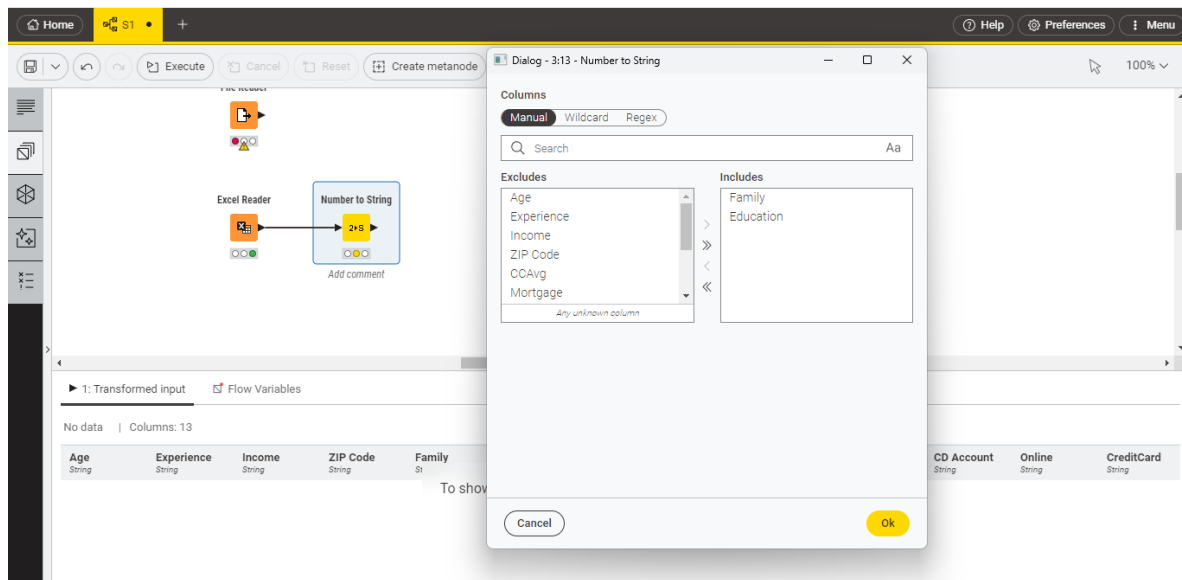


Figura 2: Number to String

Con este nodo nos aseguramos que estas variables se manejen correctamente como categóricas en el análisis.

b) Utiliza el nodo Data Explorer para una visualización y estudio básico de las variables de la base de datos. ¿Existe alguna característica que te llame la atención? ¿Por qué?

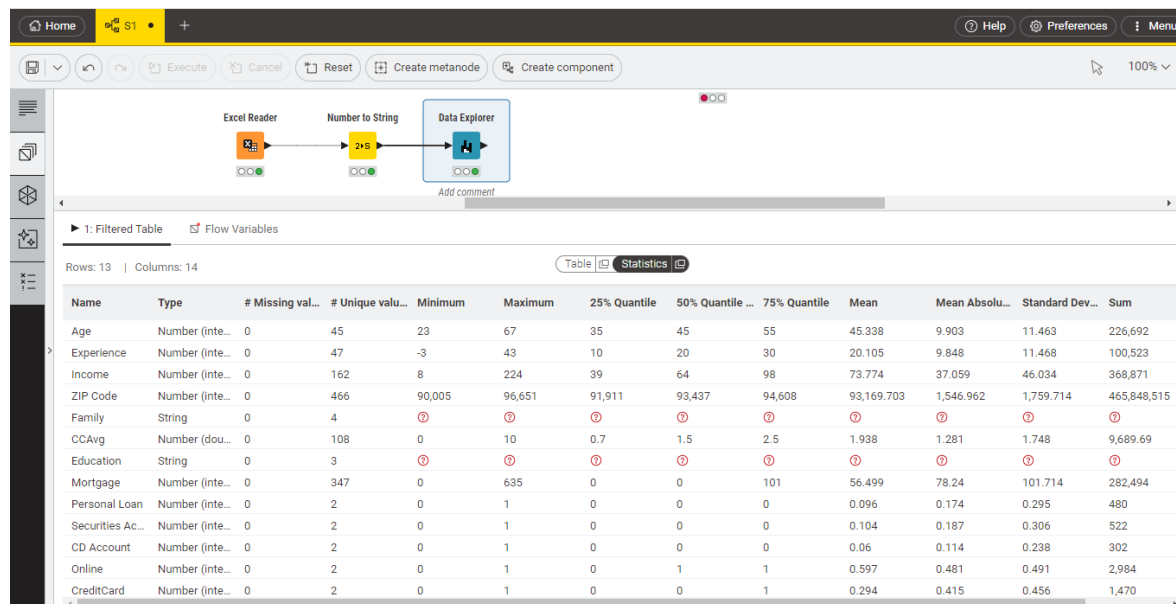


Figura 3: Data Explorer

Las variables categóricas, aunque contengan valores numéricos, deben tratarse como cualitativas para evitar cálculos inapropiados. Su análisis se realiza mejor utilizando medidas como la moda y pruebas de Chi-cuadrado para estudiar relaciones entre categorías. En el caso de las variables booleanas (0 y 1), aunque pueden ser numéricas, su análisis es más significativo usando proporciones o la frecuencia relativa. Tratar estos datos correctamente evita errores en el análisis estadístico y asegura resultados más precisos.

c) Realiza un diagrama de burbujas para analizar el número de miembros de la familia de las personas recogidas en la tabla. Dibuja un diagrama de barras de esta característica.

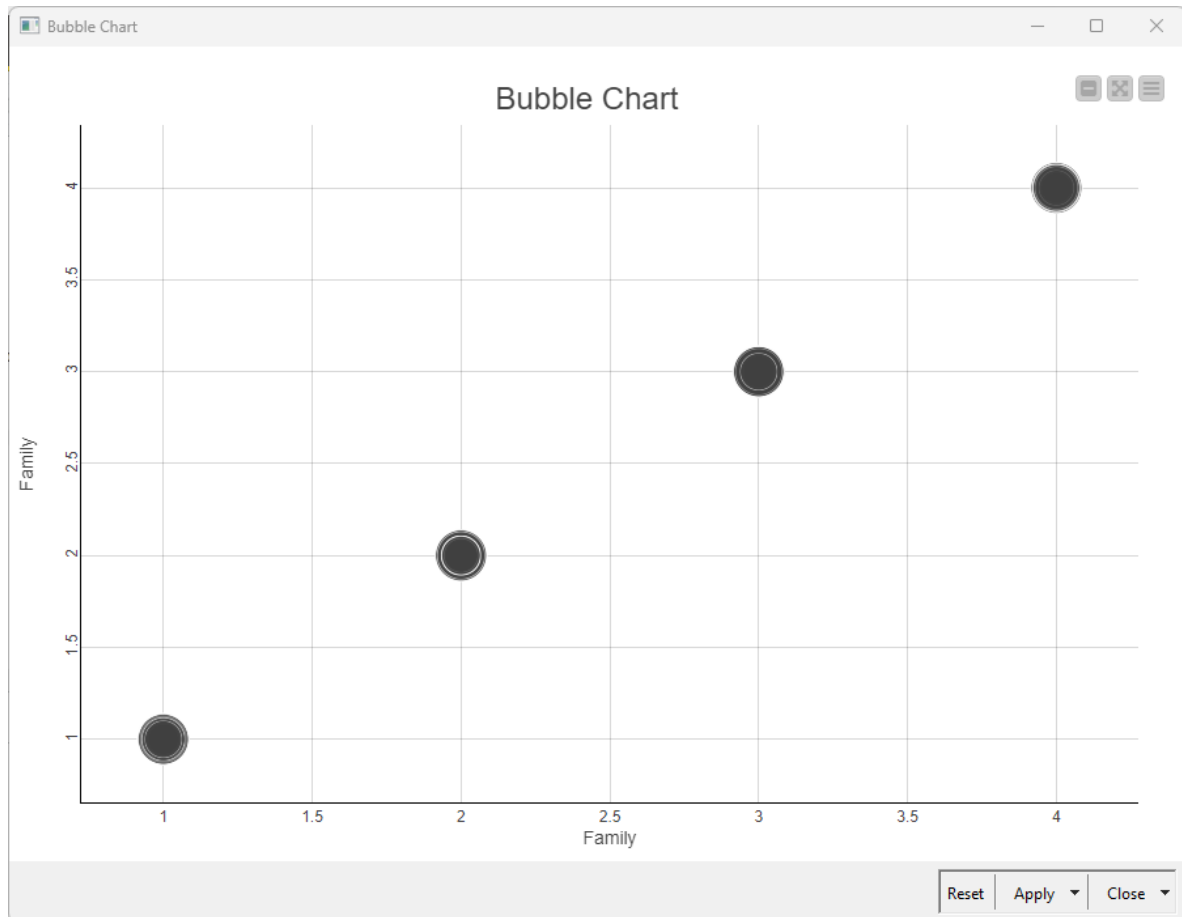


Figura 4: Bubble Chart

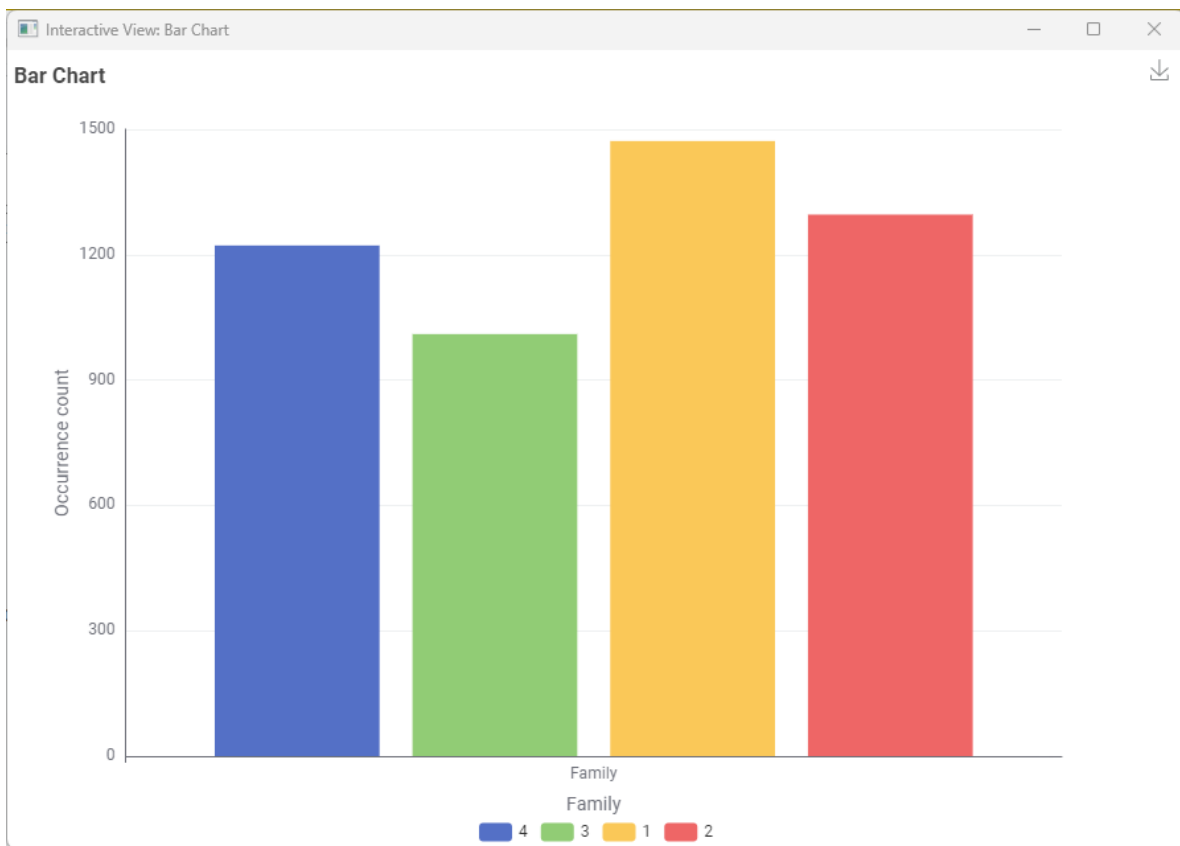


Figura 5: Bar Chart

El diagrama de burbujas no es ideal para analizar el número de miembros de la familia, ya que dificulta la interpretación precisa de frecuencias. El diagrama de barras es más adecuado porque muestra claramente la distribución de la variable categórica, permitiendo una comparación directa y fácil de las frecuencias relativas de cada categoría. Matemáticamente, el diagrama de barras es óptimo para visualizar datos discretos, ya que representa de manera clara las proporciones y facilita el uso de pruebas estadísticas como la comparación de proporciones entre grupos.



d) Realiza un diagrama de dispersión para analizar visualmente los campos edad y años de experiencia. ¿Que opinas sobre la gráfica? Estudia el coeficiente de correlación entre dichas variables.

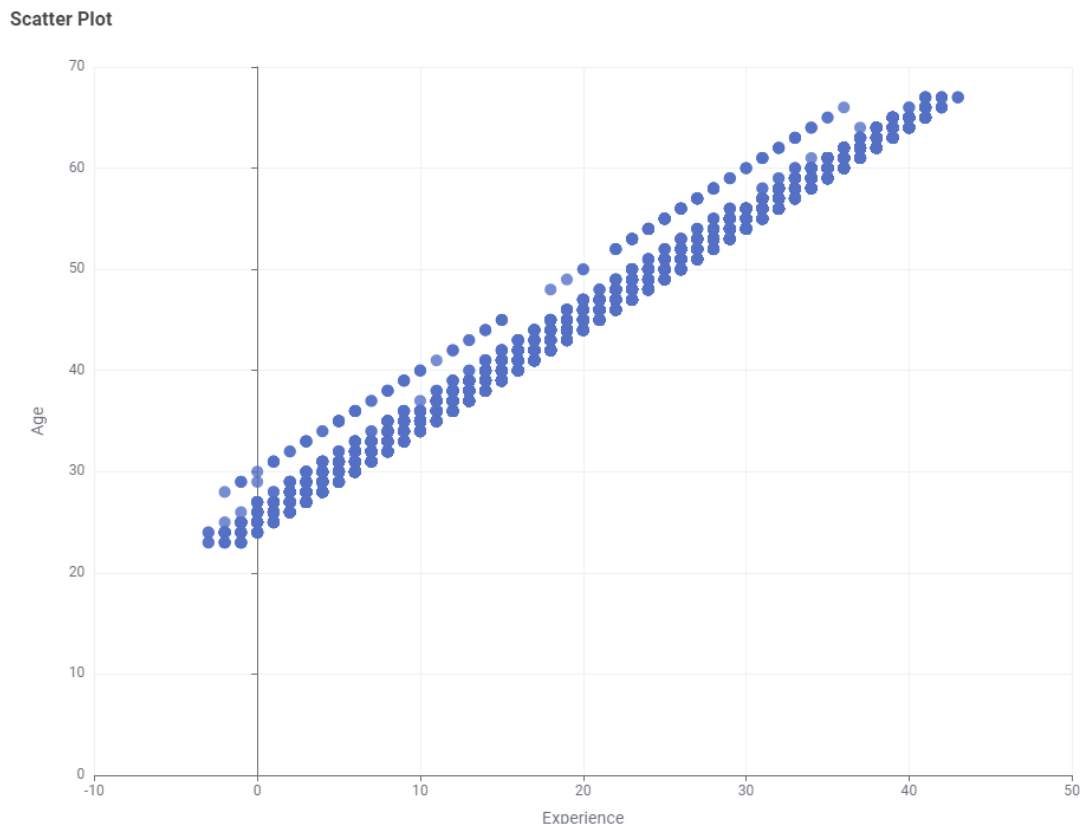


Figura 6: Scatter Plot

El diagrama de dispersión muestra una clara correlación lineal positiva entre la edad y los años de experiencia. Matemáticamente, esta relación se cuantifica mediante el coeficiente de correlación de Pearson, que mide la fuerza y dirección de la asociación lineal entre las dos variables. Un valor de Pearson cercano a 1 confirma una fuerte correlación positiva, lo cual es lógico ya que, en general, una mayor edad está asociada a más años de experiencia. Este comportamiento indica una dependencia directa y predecible entre ambas variables.

e) Estudia las matrices de dispersión del resto variables. Calcula la correlación lineal para las variables en las que eso tenga sentido. ¿Sería razonable eliminar alguna característica en base a este estudio?

Uso el nodo Scatter Plot Matrix para sacar la matrix de todos los valores.

Scatter Plot Matrix

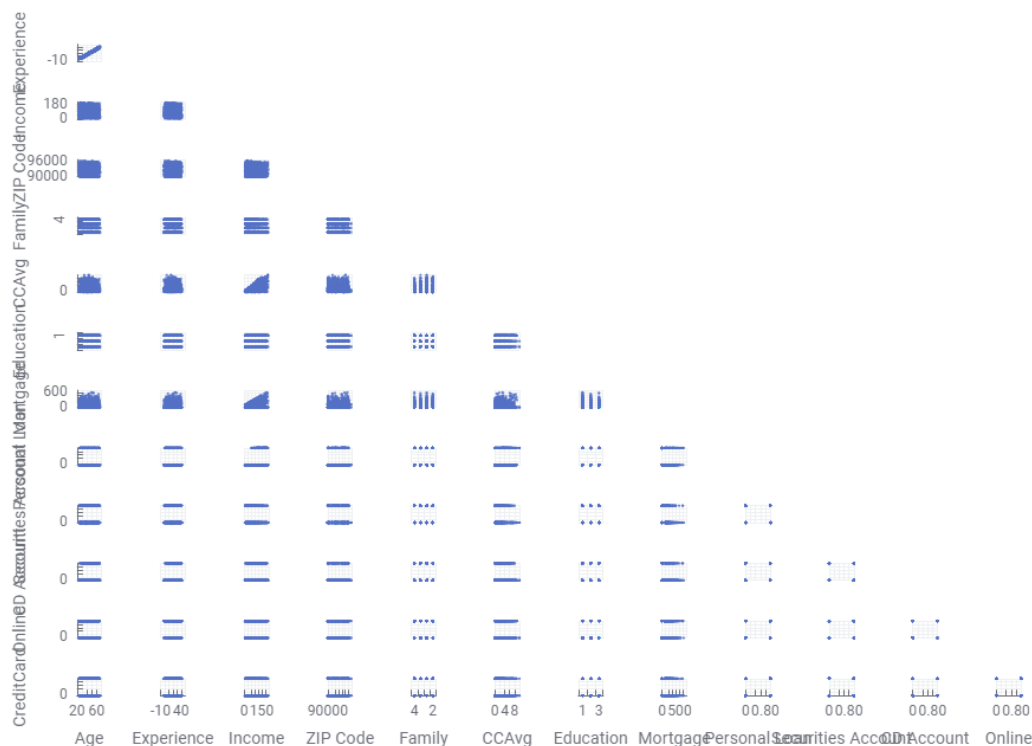


Figura 7: Scatter Plot Matrix

Para poder ver mejor la correlación entre las variables aplico el nodo Linear Correlation: En este apartado, se presenta el análisis de correlaciones entre las principales variables del conjunto de datos. Se utilizaron coeficientes de correlación de Pearson para medir la fuerza y dirección de las relaciones lineales entre las variables. A continuación se destacan las correlaciones más relevantes:

#### Edad y Experiencia:

- Correlación: 0.994
- Interpretación: Existe una correlación extremadamente alta entre la edad y la experiencia, lo que indica que a medida que aumenta la edad, también aumenta la experiencia laboral.

#### Ingresos y Gasto Promedio con Tarjeta (CCAvg):

- Correlación: 0.646
- Interpretación: Los ingresos más altos están asociados con un mayor gasto promedio en tarjetas de crédito, sugiriendo hábitos de consumo relacionados con la capacidad económica.

#### Ingresos y Hipoteca

- Correlación: 0.207
- Interpretación: Esta correlación moderada sugiere que los hogares con mayores ingresos tienden a tener hipotecas más grandes.

#### Gasto Promedio con Tarjeta (CCAvg) y Préstamo Personal

- Correlación: 0.367
- Interpretación: Se observa que aquellos con un mayor gasto promedio en tarjetas de crédito son más propensos a tener préstamos personales, lo cual puede reflejar comportamientos de deuda.

#### Educación y Tamaño de la Familia:

- Correlación: 0.130
- Interpretación: Aunque la correlación es baja, es importante considerar cómo el nivel educativo podría afectar el tamaño de la familia en el contexto socioeconómico.

#### Préstamo Personal y Cuenta de Depósito a Plazo (CD Account):

- Correlación: 0.316
- Interpretación: Indica una relación positiva, sugiriendo que las personas con préstamos personales también tienden a tener cuentas de depósito a plazo.

#### Cuenta de Depósito a Plazo (CD Account) y Tarjeta de Crédito:

- Correlación: 0.279
- Interpretación: Sugiere que aquellos con cuentas de depósito a plazo tienen cierta relación con la tenencia de tarjetas de crédito, lo que puede reflejar comportamientos financieros complementarios.

El análisis de correlaciones realizado sobre el conjunto de datos ha permitido identificar relaciones significativas entre varias variables. Las variables con alta correlación son:

1. Edad y Experiencia: La correlación de 0.994 indica una relación casi perfecta entre la edad y los años de experiencia laboral. Esto sugiere que los individuos tienden a adquirir más experiencia a medida que envejecen, lo cual es consistente con la naturaleza del desarrollo profesional a lo largo de la vida.
2. Ingresos y Gasto Promedio con Tarjeta (CCAvg): Con una correlación de 0.646, se observa que los individuos con mayores ingresos tienden a tener un mayor gasto promedio en tarjetas de crédito.
3. Gasto Promedio con Tarjeta (CCAvg) y Préstamo Personal: La correlación de 0.367 sugiere que aquellos con un mayor gasto promedio en tarjetas de crédito son más propensos a tener préstamos personales.
4. Préstamo Personal y Cuenta de Depósito a Plazo (CD Account): La correlación de 0.316 sugiere que las personas con préstamos personales también son más propensas a mantener cuentas de depósito a plazo.

f) Asigna colores a las filas en al número de componentes de la familia. A partir de esto, crea un diagrama de dispersión entre la edad y el salario donde los puntos tengan un color dependiendo del número de componentes de la familia. ¿Se puede observar algo en este diagrama?



Figura 8: Scatter Plot - f)

El análisis de dispersión muestra una tendencia leve que sugiere que las familias con un mayor número de miembros tienden a tener ingresos más bajos y una edad promedio más joven. Además, una edad promedio más joven puede reflejar la formación de familias en etapas tempranas de la vida, lo que podría estar relacionado con menores niveles de experiencia laboral y, por ende, menores ingresos.

g) Filtra la base de datos según la variable Family (nodo Row Filter). Por ejemplo, considera los clientes con familias de 4 miembros (o cualquier otro número, si lo prefieres). Realiza un estudio parecido a los puntos anteriores. ¿Se puede observar algo destacable para estos clientes?

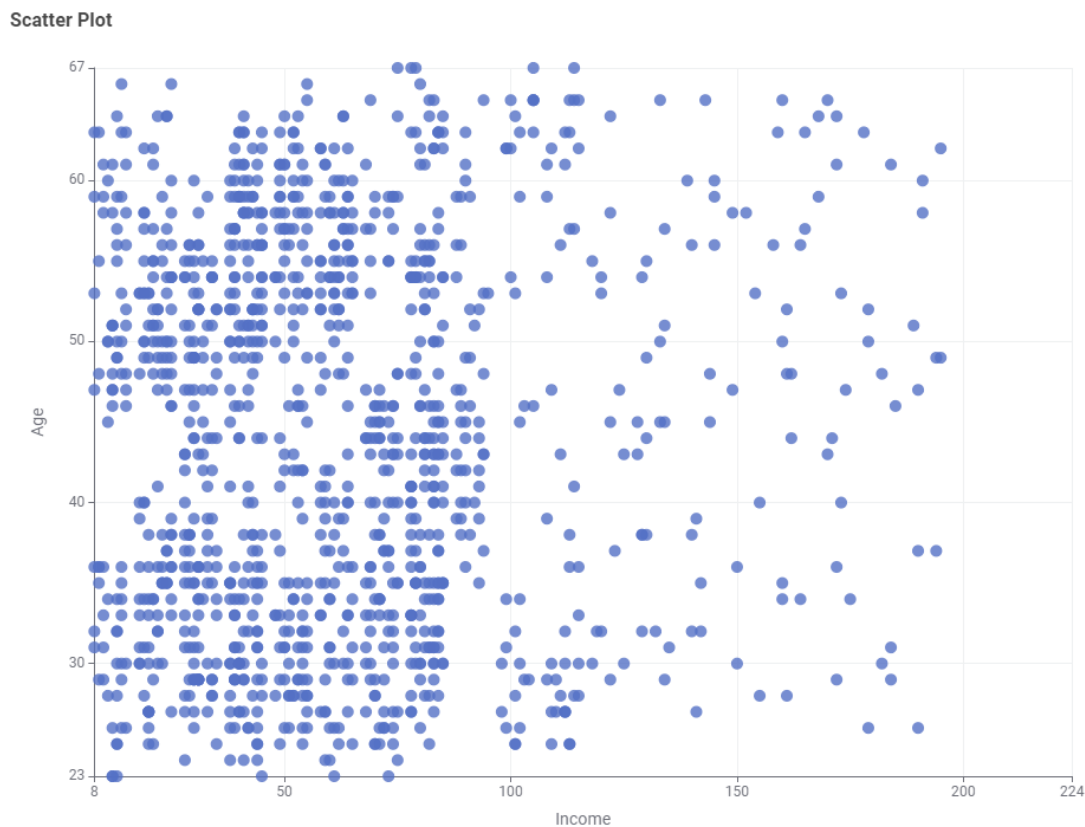


Figura 9: Scatter Plot - g

h) De la base de datos filtrada anterior, calcula diagramas circulares para las variables nominales.

En el análisis de la base de datos filtrada, se identificaron dos variables nominales: educación y número de miembros de la familia.

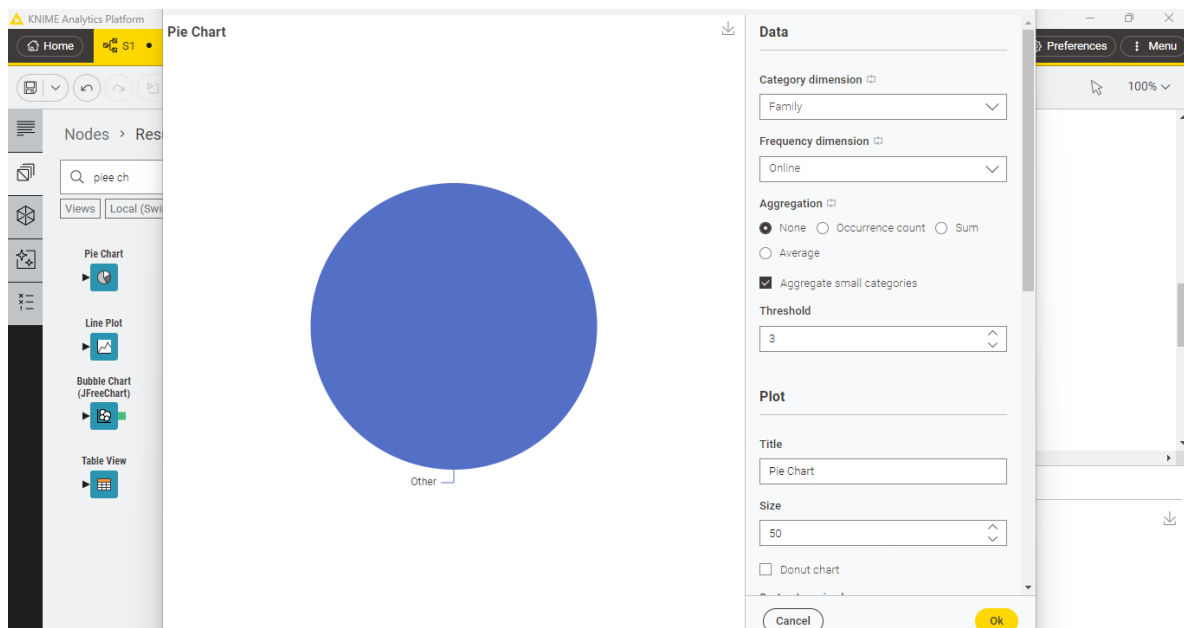


Figura 10: Pie Chart - family

Al realizar un análisis por nivel educativo en familias de cuatro miembros, se pueden extraer conclusiones útiles sobre la distribución de los salarios medios. Los resultados del diagrama circular revelan que el grupo con nivel educativo 3 presenta el mayor ingreso promedio, alcanzando 68.72, lo que lo sitúa por encima de los otros grupos. Esta diferencia sugiere que a medida que aumenta el nivel educativo, también lo hace el salario medio, indicando que una educación superior puede estar asociada con mayores oportunidades económicas.

## Pie Chart

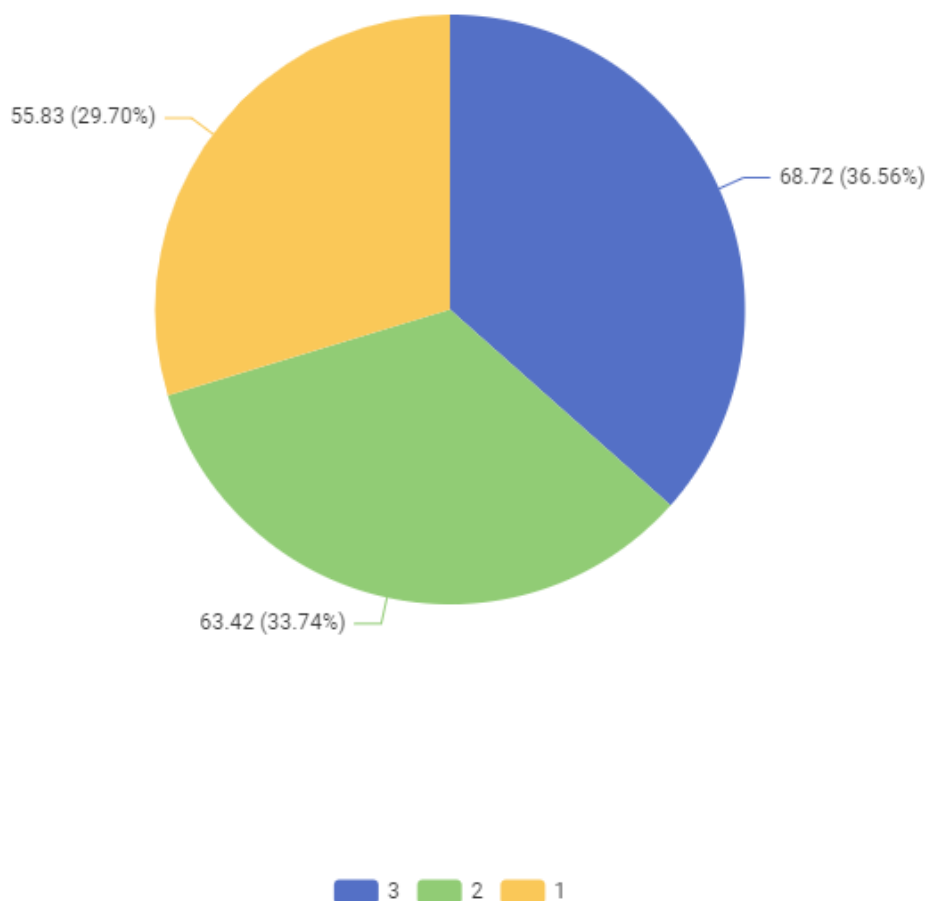


Figura 11: Pie Chart - Educacion - salarios

Al analizar el gasto promedio con tarjeta (CCAvg) en familias de cuatro miembros, se observa que el grupo con nivel educativo 3 presenta el mayor gasto promedio, con un valor medio de 1.87 y un porcentaje de 37.34. Este gasto es superior al de los niveles educativos 1 y 2, que tienen valores de 1.52 y 1.62, respectivamente, con porcentajes de 30.30 y 32.36.

Estos resultados indican que a medida que aumenta el nivel educativo, también lo hace el gasto promedio con tarjeta. Esto puede sugerir que las familias con un mayor nivel educativo tienen mayor acceso a recursos y, por ende, tienden a gastar más en bienes y servicios, reflejando posiblemente un mayor poder adquisitivo y hábitos de consumo más elevados.

## Pie Chart

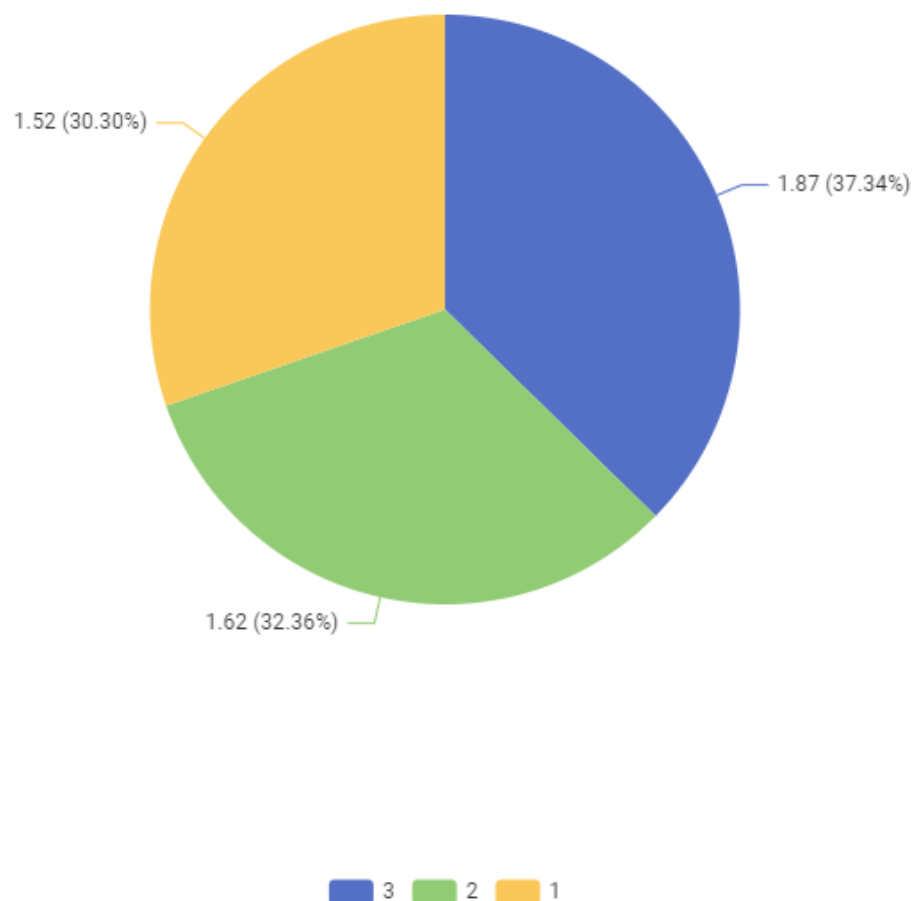


Figura 12: Pie Chart - Educacion y CCAvg

Al analizar los préstamos personales (Personal Loan) en familias de cuatro miembros, se observa que el grupo con nivel educativo 3 tiene la mayor proporción de préstamos personales, con un valor medio de 0.12 y un porcentaje de 36.62. Esto es superior a los niveles educativos 1 y 2, que presentan valores de 0.11 y 0.10, con porcentajes de 29.45 y 33.93, respectivamente.

Estos resultados sugieren que las familias con un nivel educativo más alto son más propensas a tener préstamos personales, lo que puede reflejar una mayor confianza en la capacidad de pago y un mejor acceso a oportunidades de financiamiento. Este patrón puede indicar que un mayor nivel educativo está asociado con una mejor comprensión y gestión de productos financieros.



## Pie Chart

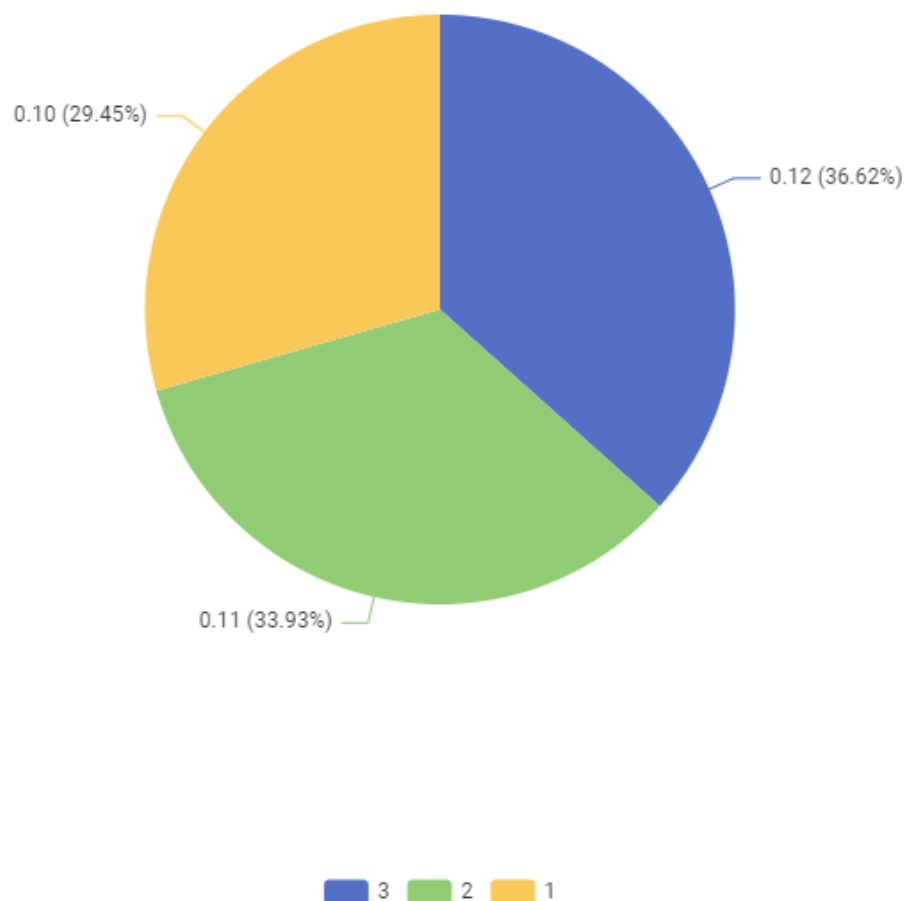


Figura 13: Pie Chart - Education y personal loan

Al analizar las cuentas de depósito a plazo (CD Account) en familias de cuatro miembros, se observa que el grupo con nivel educativo 3 presenta la mayor proporción, con un valor medio de 0.07 y un porcentaje de 42.5. Esto es superior a los niveles educativos 1 y 2, que tienen valores de 0.05 y porcentajes de 27.57 y 29.93, respectivamente.

Estos resultados indican que las familias con un nivel educativo más alto son más propensas a tener cuentas de depósito a plazo, lo que puede reflejar una mayor capacidad para ahorrar y una gestión financiera más sólida. Esta tendencia sugiere que un mayor nivel educativo está asociado con un mejor manejo de los ahorros y la inversión en productos financieros.

## Pie Chart

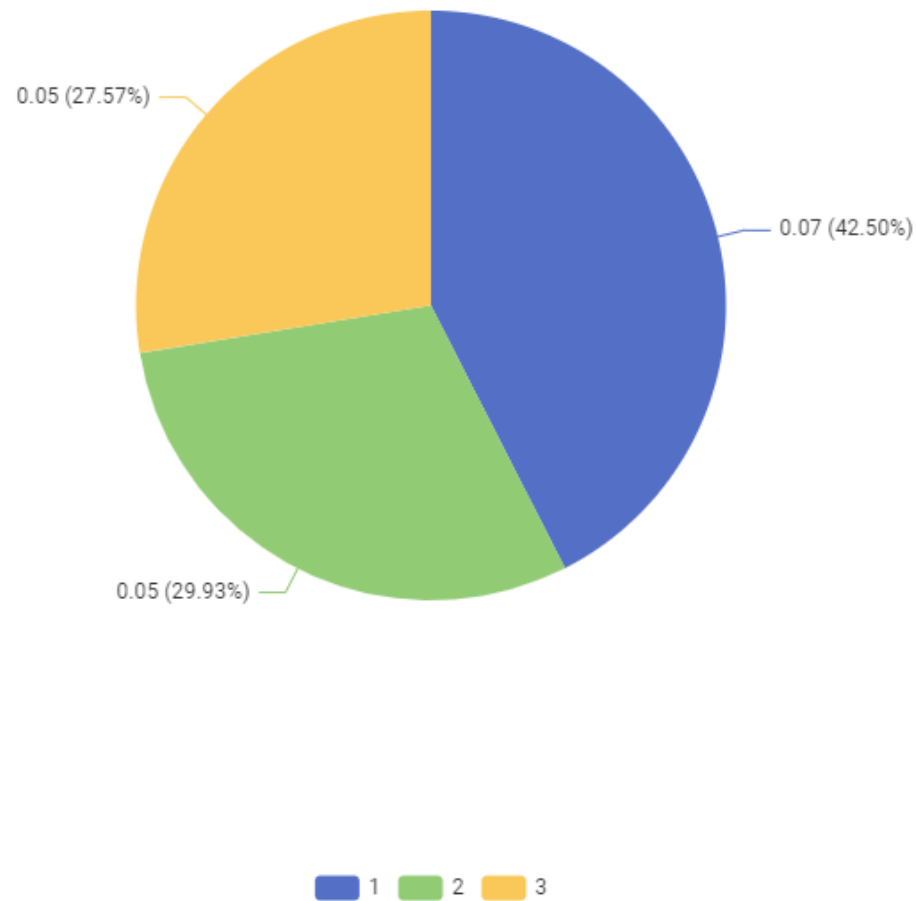


Figura 14: Pie Chart - Education y CD Account

## Ejercicio 2

(4 puntos) Descarga el fichero NBA.xlsx desde la página de la asignatura de Google Classroom. Contiene datos sobre los jugadores de la NBA que han jugado los playoffs durante la temporada 2021-2022.



a) El nombre de las variables es demasiado largo. Cámbialos por otros más cortos utilizando el nodo Column Rename.

The screenshot shows a dialog box titled "Dialog - 3:28 - Column Renamer". It contains a list of columns being renamed, with each entry having a "Column" field, a "New name" field, and a trash icon. The columns are:

Column	New name
APGAssistsAssists per game.	Assists per game
RPGReboundsRebounds per game.	Rebounds per game
PPGPointsPoints per game.	Points per game
TS%True Shooting PercentageTrue	True Shooting Percentage
eFG%Effective Shooting Percentage	Effective Shooting Percentage
USG%Usage RateUsage rate, a.k.a.,	Usage Rate
TO%Turnover RateA metric that esti	Turnover Rate

At the bottom of the dialog, there is an "Add column" button with a plus icon, a "Cancel" button, and an "Ok and Execute" button.

Figura 16: Column Renamer

b) Calcula las medidas básicas utilizando el nodo Statistics. ¿Existe algún dato que te llame la atención?



Figura 17: Statistics



Figura 18: Statistics 2

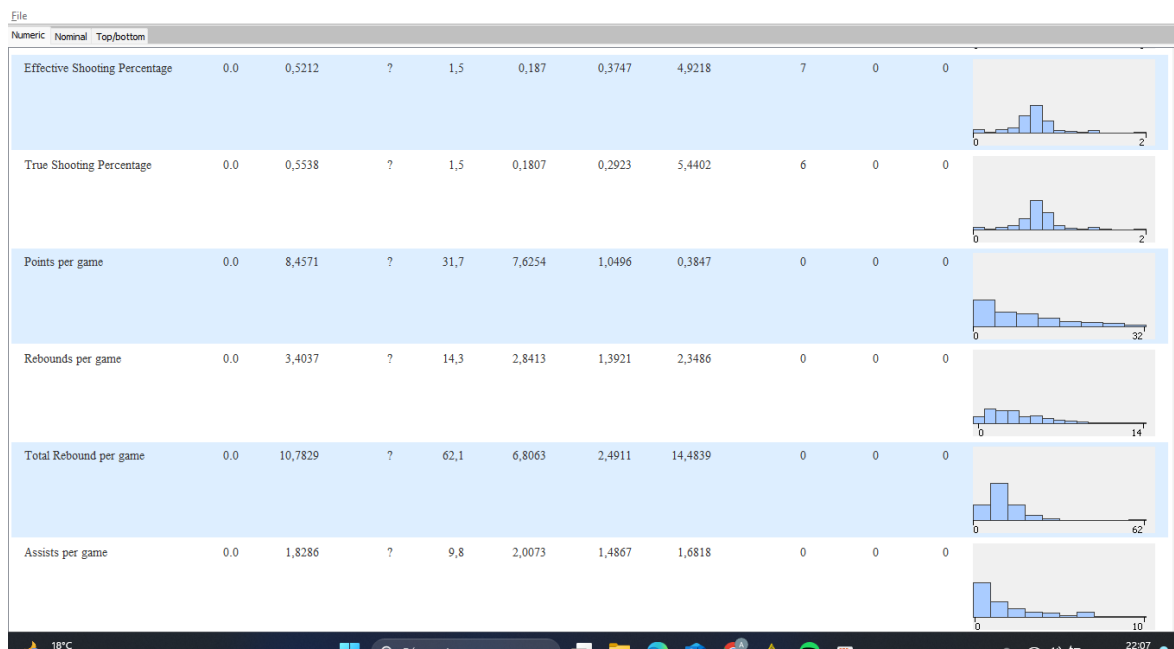


Figura 19: Statistics 3

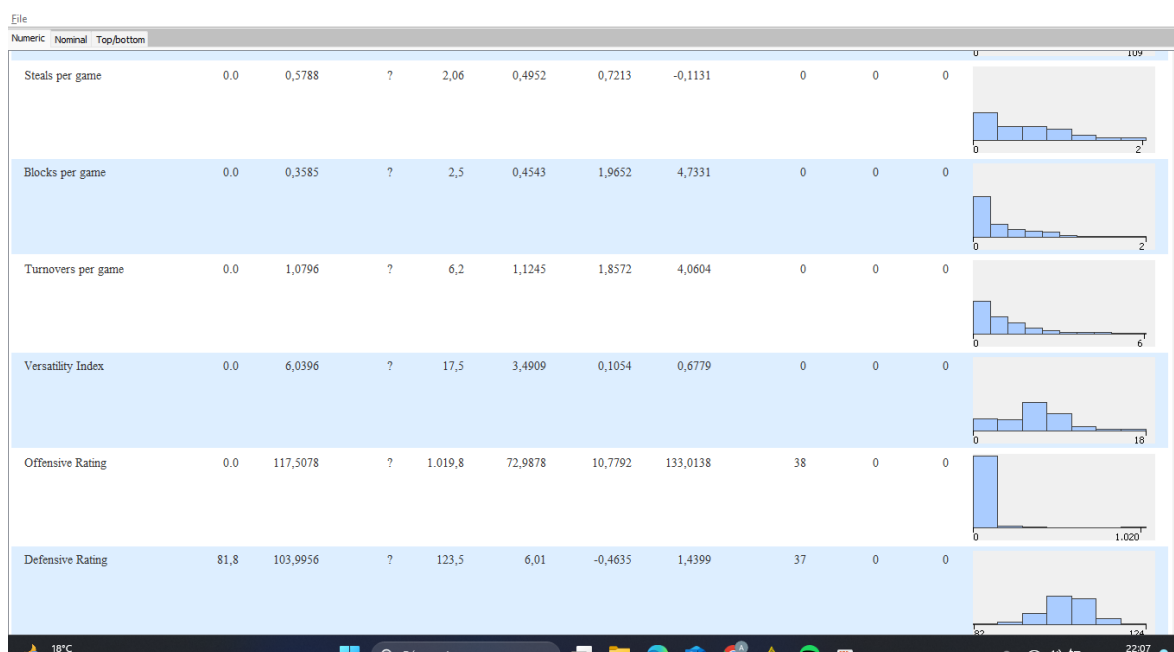


Figura 20: Statistics 4

- **Asimetría y Curtosis en Uso de Posesiones y Tasa de Pérdidas:** Las variables como el *Usage Rate* y *Turnover Rate* muestran una alta asimetría positiva y curtosis elevada, indicando distribuciones sesgadas con valores extremos.
- **Desviaciones significativas en Offensive Rating:** El *Offensive Rating* tiene una desviación estándar muy alta y un valor máximo de 1019.8, lo que sugiere la presencia de outliers significativos en el rendimiento ofensivo.

- **Distribución de Minutos Jugados y Porcentaje de Minutos:** El *Minutes Percentage* presenta un valor máximo del 91.7%, con una asimetría negativa, lo que indica que la mayoría de los jugadores tienen un porcentaje de minutos jugados por debajo del promedio.
- **Distribución de Estadísticas de Tiros (FT %, 2P %, 3P %):** Los porcentajes de tiro muestran distribuciones aproximadamente simétricas, sugiriendo una distribución equilibrada en las habilidades de tiro entre los jugadores.
- **Discrepancia en Offensive Rating y Defensive Rating:** Se observa una mayor consistencia en el *Defensive Rating* en comparación con el *Offensive Rating*, lo cual está reflejado en la menor dispersión de esta variable.

c) Realiza diagramas de dispersión de las variables numéricas de tipo real (double). Al observar las gráficas, ¿consideras que algunas variables están relacionadas? ¿Cuáles? ¿Existe alguna explicación razonable para esto?

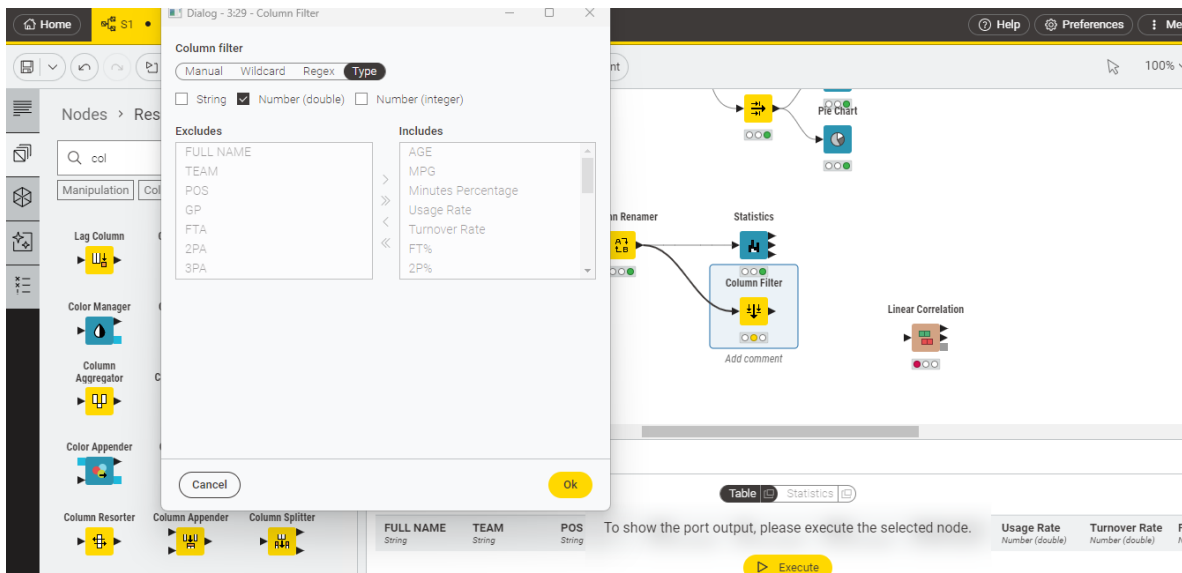


Figura 21: Column Filter

Al observar la matriz de gráficos de dispersión, las gráficas no se distinguen con claridad debido al tamaño reducido de cada gráfico individual y la superposición de puntos. Esto hace que sea complicado identificar visualmente las correlaciones entre las variables sin una inspección más detallada.

Y aunque la visualización de las gráficas no permite una fácil interpretación a simple vista, al profundizar en los datos se pueden identificar algunas relaciones clave que destacan el rendimiento integral de los jugadores y la importancia del tiempo de juego.

d) Calcula matriz de correlación de las variables numéricas utilizando el nodo Linear correlation. Compara con el punto anterior.

► 1: Correlation measure   ► 2: Correlation matrix   ■ 3: Correlation model   📄 Flow Variables

Rows: 210 | Columns: 5   **Table** 📄 Statistics 📄 🔍

<input type="checkbox"/>	#	RowID	First column name String	Second column name String	Correlation value Number (double)	p value Number (double)	Degrees of freedom Number (integer)
<input type="checkbox"/>	1	Row0	AGE	MPG	0.177	0.009	215
<input type="checkbox"/>	2	Row1	AGE	Minutes Percentage	0.177	0.009	215
<input type="checkbox"/>	3	Row2	AGE	Usage Rate	-0.046	0.497	215
<input type="checkbox"/>	4	Row3	AGE	Turnover Rate	0.116	0.093	210
<input type="checkbox"/>	5	Row4	AGE	FT%	0.126	0.065	215
<input type="checkbox"/>	6	Row5	AGE	2P%	0.002	0.977	215
<input type="checkbox"/>	7	Row6	AGE	3P%	0.006	0.929	215
<input type="checkbox"/>	8	Row7	AGE	Effective Shooting Percentage	-0.047	0.497	208
<input type="checkbox"/>	9	Row8	AGE	True Shooting Percentage	-0.041	0.555	209
<input type="checkbox"/>	10	Row9	AGE	Points per game	0.067	0.323	215
<input type="checkbox"/>	11	Row10	AGE	Rebounds per game	0.159	0.019	215
<input type="checkbox"/>	12	Row11	AGE	Total Rebound per game	0.135	0.048	215
<input type="checkbox"/>	13	Row12	AGE	Assists per game	0.165	0.015	215
<input type="checkbox"/>	14	Row13	AGE	Assist Percentage	0.016	0.817	215
<input type="checkbox"/>	15	Row14	AGE	Steals per game	0.091	0.18	215
<input type="checkbox"/>	16	Row15	AGE	Blocks per game	0.049	0.476	215
<input type="checkbox"/>	17	Row16	AGE	Turnovers per game	0.007	0.175	215

Figura 22: Linear Correlation

La matriz de correlación generada mediante el nodo de correlación lineal proporciona una comparación con las correlaciones observadas anteriormente. Aquí se presentan las relaciones de correlación entre las variables numéricas, destacando las siguientes observaciones:

- AGE y otras variables: Las correlaciones de AGE con otras variables como MPG, Minutes Percentage y Usage Rate se mantienen relativamente bajas, con valores de correlación alrededor de  $\pm 0.17$ . Esto sugiere que la edad no tiene una fuerte relación lineal con estas métricas.
- MPG y Minutes Percentage: Como era de esperar, la correlación entre MPG y Minutes Percentage es extremadamente alta (0.999), lo cual es lógico debido a que ambas métricas están estrechamente relacionadas y reflejan la cantidad de tiempo que un jugador pasa en el campo.
- FT %, 3P %, y Offensive Rating: Las correlaciones con métricas de eficiencia, como FT % y 3P %, varían dependiendo de la métrica con la que se comparen. Por ejemplo, MPG muestra una correlación moderada con FT % (0.511) y 3P % (0.277), mientras que la relación de Offensive Rating con otras métricas es menos significativa.
- Variables relacionadas con el rendimiento: Variables como Points per game, Rebounds per game y Assists per game muestran una alta correlación con MPG, lo que indica que cuanto más tiempo pasa un jugador en el campo, mejor es su rendimiento en estas áreas.
- Significancia estadística: Es importante tener en cuenta los valores p asociados a cada correlación para evaluar su significancia. Muchas de las correlaciones que tienen un valor p inferior a 0.05 se consideran estadísticamente significativas, lo que indica que hay una relación confiable entre las variables.



e) Realiza un diagrama de barras de la columna POS (posición del jugador).

Bar Chart

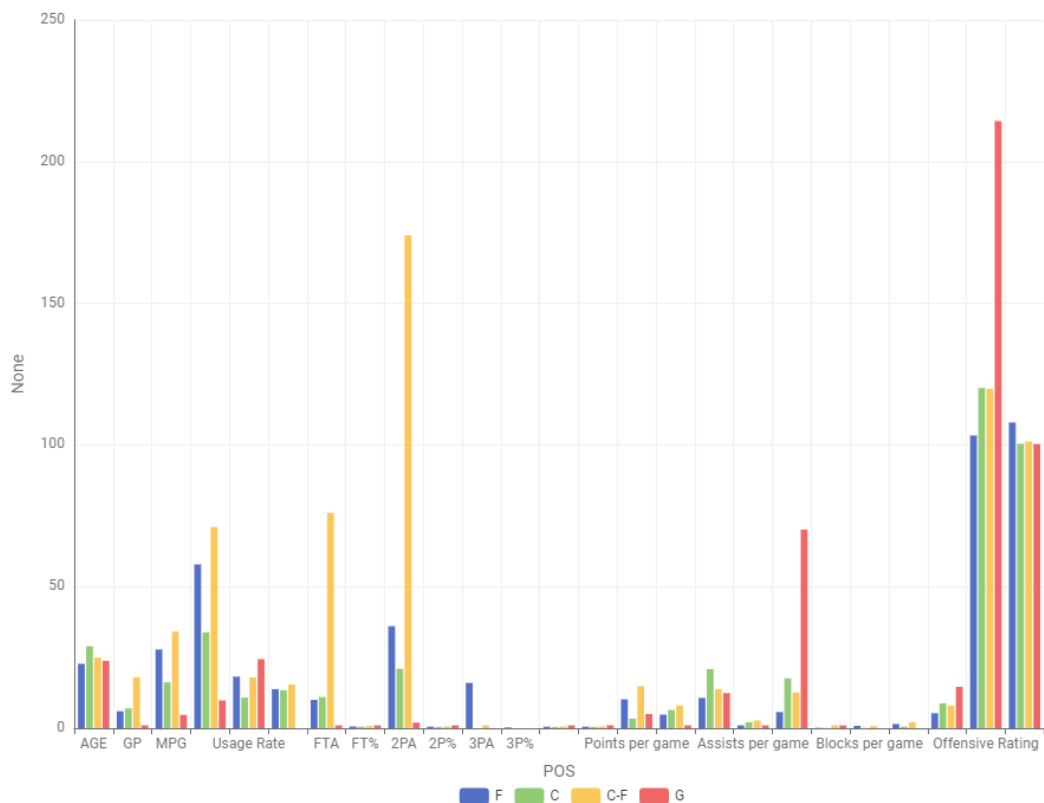


Figura 23: Bar Chart

El diagrama de barras que se muestra compara diferentes métricas de rendimiento de jugadores según sus posiciones en la cancha (POS), representadas por categorías como F (Forward), C (Center), C-F (Center-Forward) y G (Guard). Cada barra muestra la contribución de cada posición para las métricas como AGE, MPG, Usage Rate, FT %, 2PA, 3P %, entre otras.

Para entender mejor estas categorías, podemos observar cómo cada posición se desempeña en las métricas específicas. Por ejemplo, es notable cómo los C-F dominan en el campo de 2PA (intentos de tiro de 2 puntos), mientras que los Guards (G) tienden a tener un desempeño más significativo en asistencias por juego.

f) Realiza diagramas de cajas de las variables numéricas. ¿Existen jugadores que consideras que sobresalen del resto en alguna característica (outliers)?

Box Plot

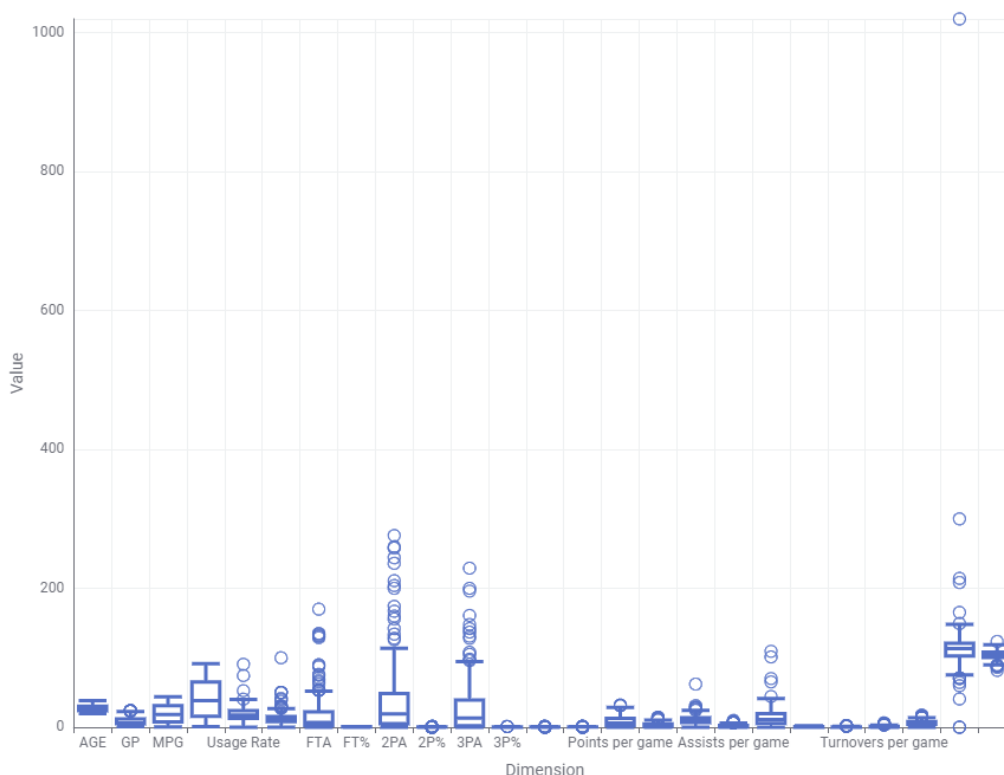


Figura 24: Box Plot

Es posible identificar a jugadores que sobresalen del resto mediante la observación de los valores atípicos (outliers). Los outliers se muestran como puntos fuera del rango de las cajas y representan a jugadores cuyas características se desvían significativamente del resto.

- FT % (porcentaje de tiros libres) y 2PA (intentos de tiros de dos puntos) muestran varios outliers, lo que indica que algunos jugadores tienen valores extremadamente altos en estas categorías en comparación con la mayoría de los jugadores.
- Usage Rate y Points per game (puntos por partido) también parecen tener algunos valores fuera del rango esperado, lo que podría indicar jugadores con un rol ofensivo muy dominante.
- Turnovers per game (pérdidas de balón por partido) muestra claramente algunos outliers, lo que sugiere que algunos jugadores cometen significativamente más pérdidas de balón que el promedio.

g) Calcula un diagrama circular de la columna GP (partidos jugados).

### Pie Chart

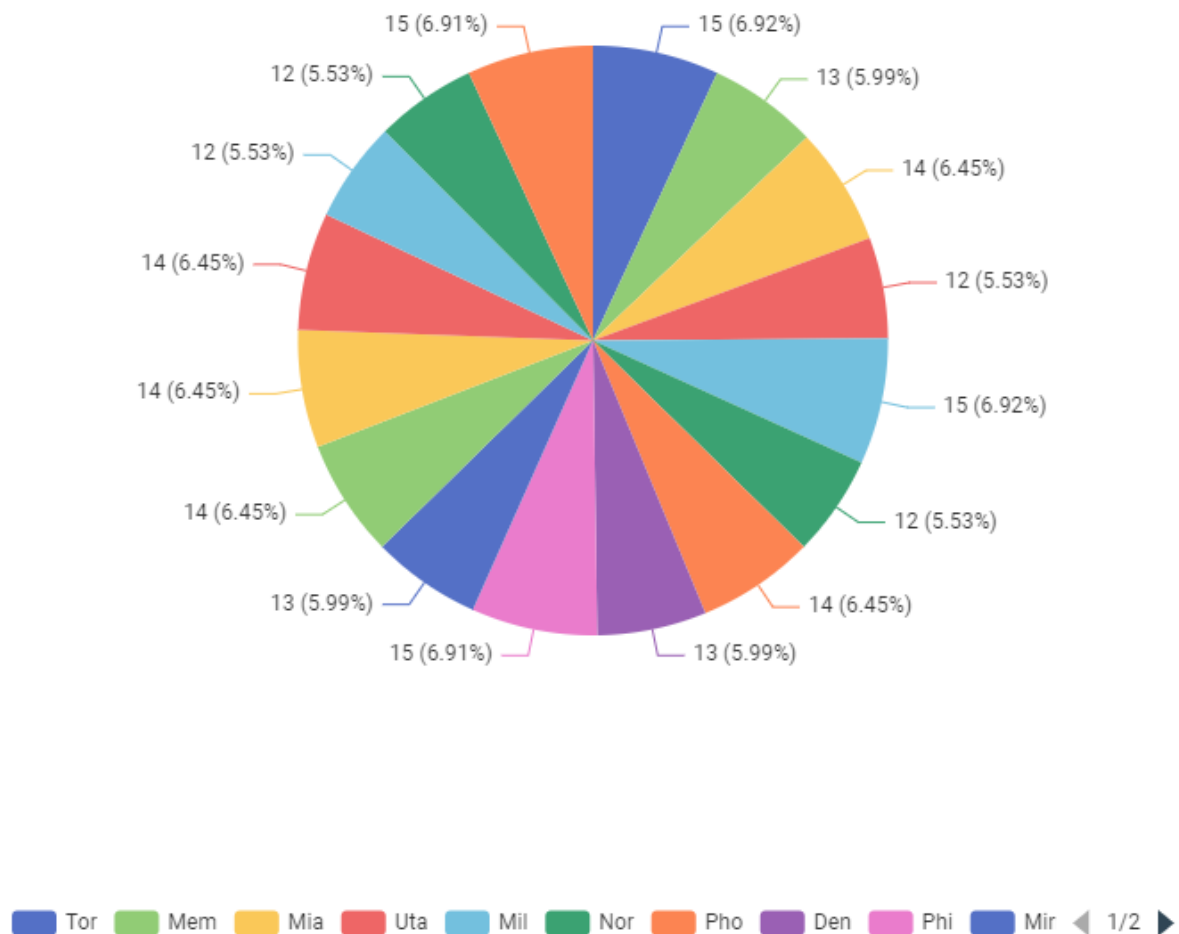


Figura 25: Pie Chart

Podemos observar que todos los equipos han jugado la misma cantidad de partidos, Este tipo de gráfico no revelará mucha variación ya que todos los segmentos del círculo serán iguales.

## Pie Chart

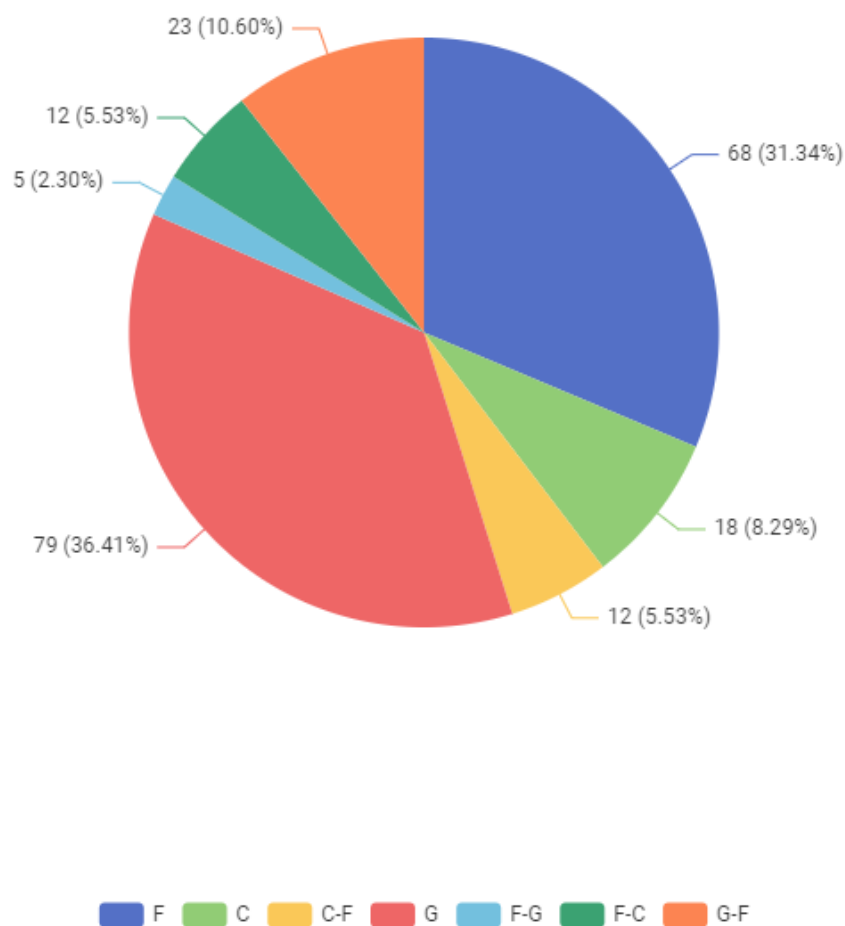


Figura 26: Pie Chart - Pos

En este caso, hay una diferencia en el número de partidos jugados según las posiciones de los jugadores. El análisis muestra que:

1. Las posiciones G (Guardias) y F (Aleros) son las que más partidos han jugado, lo que sugiere que son roles muy utilizados en los equipos.
2. Les siguen las posiciones mixtas como G-F y C (Centro), que también tienen una cantidad significativa de partidos jugados.
3. Finalmente, las posiciones como C-F, F-C, y F-G son las que menos partidos han jugado, lo que indica que son menos frecuentes o menos utilizadas.



h) Selecciona un equipo y filtra la tabla para obtener los jugadores que pertenecen a dicho equipo, usando el nodo Row Filter. Selecciona 5 características numéricas y filtra la tabla para quedarse únicamente con esas columnas (además de la columna con el nombre). Con la tabla resultante dibuja un diagrama de coordenadas paralelas (dibuja la líneas paralelas en colores a partir de la columna del nombre). ¿Es posible observar algún jugador con unas características diferentes del resto? Prueba con otros equipos y otras características, si no ha sido posible.

Dialog - 3:34 - Row Filter

### Filter

Criterion 1

Filter column: TEAM

Operator: Equals

Case matching: Case sensitive (selected), Case insensitive

Value: Tor


+ Add criterion

Output



Cancel Ok



Figura 27: Row Filter

En este caso voy a seleccionar los jugadores del equipo Tor:

► 1: Included Rows 

Rows: 15 | Columns: 28 

Table  Statistics 

<input type="checkbox"/>	#	RowID	FULL NA... <small>String</small>	TEAM <small>String</small>	POS <small>String</small>	AGE <small>Number (dou...)</small>	GP <small>Number (inte...)</small>	MPG <small>Number (dou...)</small>	Minutes ... <small>Number (dou...)</small>	Usage Rate <small>Number (dou...)</small>	Turnover ... <small>Number (dou...)</small>	FTA <small>Number (inte...)</small>	FT% <small>Number (dou...)</small>	2PA <small>Num...</small>
<input type="checkbox"/>	1	?0	Precious Achi...	Tor	F	22.74	6	27.8	57.8	18.2	13.8	10	0.6	36
<input type="checkbox"/>	2	?9	OG Anunoby	Tor	F	24.92	6	36.1	75.2	21.2	10.1	16	0.75	41
<input type="checkbox"/>	3	?12	Dalano Banton	Tor	F	22.61	4	1.9	4	34.4	17.7	6	0.5	2
<input type="checkbox"/>	4	?13	Scottie Barnes	Tor	F	20.88	4	33.3	69.3	18.9	9.3	16	0.813	30
<input type="checkbox"/>	5	?19	Khem Birch	Tor	C	29.72	6	10.5	21.9	13.9	15.8	0	0	12
<input type="checkbox"/>	6	?24	Isaac Bonga	Tor	G	22.61	1	2.8	5.8	14.6	0	2	1	0
<input type="checkbox"/>	7	?26	Chris Boucher	Tor	F-C	29.43	6	21.6	45.1	17.3	4.1	10	0.9	27
<input type="checkbox"/>	8	?29	Armoni Brooks	Tor	G	24.03	4	1.9	4	24.4	0	0	0	0
<input type="checkbox"/>	9	?72	Malachi Flynn	Tor	G	24.1	6	6.1	12.6	10.2	12.5	0	0	4
<input type="checkbox"/>	10	?149	Svi Mykhailiuk	Tor	G-F	25.02	3	1.6	3.4	23.1	0	1	1	0
<input type="checkbox"/>	11	?179	Pascal Siakam	Tor	F	28.21	6	43.4	90.4	24.2	9.6	36	0.861	90
<input type="checkbox"/>	12	?194	Gary Trent Jr.	Tor	G-F	23.41	6	33.2	69.2	21.7	3.2	19	0.895	43
<input type="checkbox"/>	13	?199	Fred VanVleet	Tor	G	28.31	4	35	72.9	20.1	6.6	6	0.833	18
<input type="checkbox"/>	14	?204	Yuta Watanabe	Tor	G-F	27.68	4	2.6	5.4	31.5	14.3	0	0	5
<input type="checkbox"/>	15	?214	Thaddeus Yo...	Tor	F	33.99	6	14.5	30.2	12.2	13.2	4	0.25	11

Figura 28: Jugadores Tor

Para este análisis voy a intentar seleccionar las características numéricas más interesantes para el análisis de los jugadores en un diagrama de coordenadas paralelas, para ello voy a seleccionar aquellas que suelen ser relevantes para el rendimiento y el perfil de un jugador.

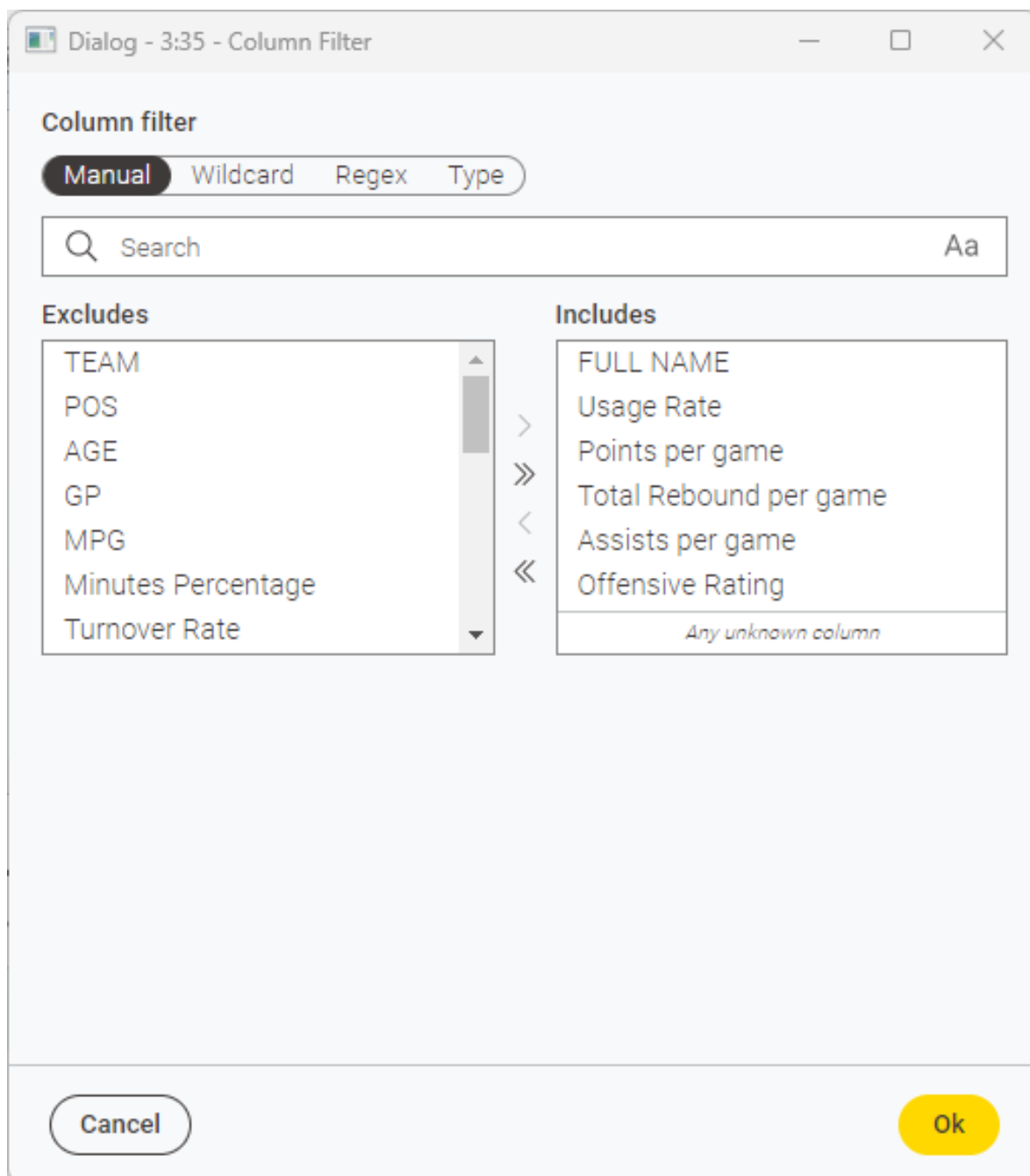


Figura 29: Extraer Caracteristicas



## Parallel Coordinates Plot

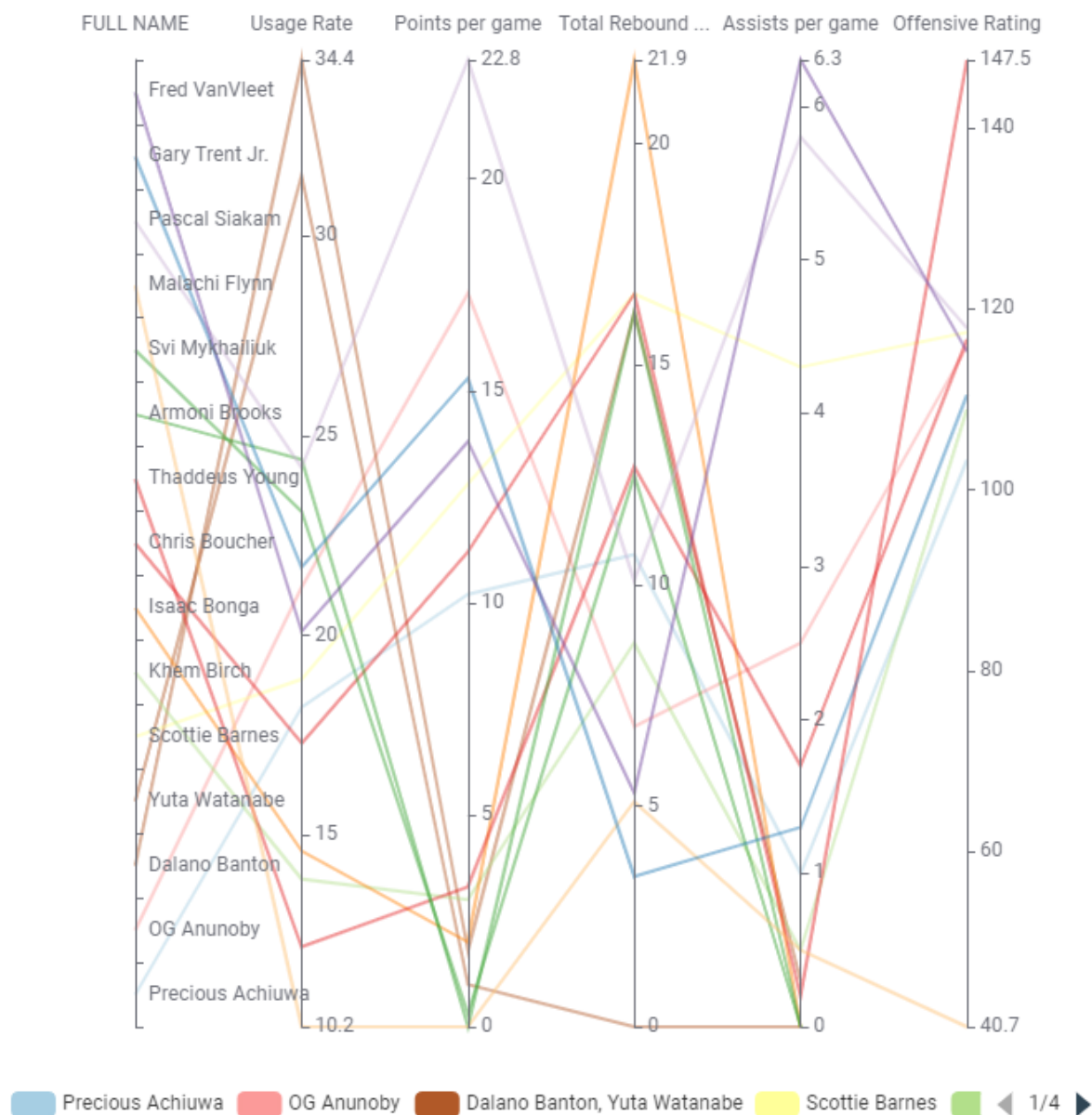


Figura 30: Parallel Coordinates Plot

### Diferenciación de Jugadores

En el gráfico se puede ver que algunos jugadores tienen valores extremos en ciertas características que los diferencian claramente del resto. Por ejemplo, un jugador con un alto Usage Rate y Points per game podría sobresalir como uno de los líderes ofensivos del equipo. Jugadores con valores elevados en Assists per game y Offensive Rating sugieren un papel más orientado a la creación de juego y a la eficiencia ofensiva.

### Variación en Estadísticas

Hay una variación significativa en algunas métricas como Total Rebounds y Assists per game, lo cual

indica que los roles dentro del equipo están bien definidos. Algunos jugadores se enfocan más en el reboteo, mientras que otros en asistir.

### Comparación de Jugadores

Un jugador como Fred VanVleet tiene valores altos en varias categorías importantes como Usage Rate, Points per game, y Assists per game, lo que lo destaca claramente como un jugador clave en múltiples facetas del juego. Otros jugadores, como Precious Achiuwa, se destacan más en características específicas como Total Rebounds, indicando una especialización en el reboteo.

**i) Si tuvieras que dividir los jugadores en varios grupos, basándote en los estudios realizados, ¿cómo lo harías? ¿En cuántos grupos? ¿Cómo denominarías/describirías a cada grupo?**

Basándome en los estudios realizados, dividiría a los jugadores en cuatro grupos principales basados en sus habilidades, roles y estadísticas observadas. Estos grupos se centran en las características de rendimiento más destacadas de los jugadores y cómo contribuyen a su equipo.

- **Anotadores Principales:** Estos jugadores son la fuerza principal de anotación del equipo. Se destacan en métricas como puntos por partido (Points per game) y tasa de uso (Usage Rate), lo que indica que el equipo depende de ellos para liderar la ofensiva y generar puntos.
- **Jugadores de Control y Creación:** Este grupo se compone de jugadores que se destacan en la creación de juego y la distribución del balón. Son aquellos con un alto número de asistencias por partido (Assists per game) y un control sobresaliente en la dirección del juego, facilitando oportunidades para sus compañeros.
- **Defensores y Especialistas en Rebotear:** Estos jugadores se especializan en defensa y rebotes. Sus métricas más destacadas incluyen rebotes totales (Total Rebounds) y estadísticas defensivas, mostrando su fortaleza en la captura de balones y en dificultar las oportunidades ofensivas del rival.
- **Jugadores Versátiles:** Los jugadores en este grupo son aquellos que tienen un rendimiento equilibrado en múltiples áreas del juego. Contribuyen tanto en ataque como en defensa, mostrando consistencia en puntos, rebotes, y asistencias. Son los jugadores que pueden desempeñar varias funciones según lo necesite el equipo.

Cuatro grupos son adecuados para capturar las diferencias clave entre los jugadores, permitiendo una clasificación clara que refleje sus roles más importantes en el equipo. Esta agrupación no solo considera la posición en la que juegan, sino también las estadísticas específicas que destacan en cada uno, lo que proporciona una visión más precisa de su contribución.