

SAMSUNG

Samsung Innovation Campus

| C&P Course

Sectores de la economía mejor pagados en República Dominicana

Departamento de Gestión de Sectores Económicos
(DGSE)

C&P Course

Sectores de la economía mejor pagados en República Dominicana

Proyecto Final Samsung Innovation Campus

En este proyecto buscamos analizar data obtenida de nóminas de instituciones públicas de los principales sectores de nuestro país, República Dominicana. Estos son Educación, Salud y Bancario. Dentro de estas se encuentran puntos clave para el análisis como lo son género, cargo, deducciones y sueldo neto. El objetivo de este proyecto es analizar e investigar qué sector de nuestro país tiene la mejor paga.

C&P Course

Codificación

Paso 1

- Importación de las librerías que se utilizarán para manejar la data

Librerías

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Codificación

Paso 2

- Cargar los datos. Los archivos con los que se va a trabajar.

Carga de datos

```
educacion = pd.read_excel(r"data\MINERD.xlsx")
salud = pd.read_excel(r"data\servicio_nacional_de_salud.xlsx")
bancos = pd.read_excel(r"data\superintendencia_de_bancos.xlsx")
```

Codificación

Paso 3

- Limpiar la data. En este paso se realiza la limpieza de la data para tener un mejor control sobre la cantidad con la que trabajaremos.

Limpieza

```
: columnas = ['genero', 'cargo', 'sueldo_bruto', 'sueldo_neto', 'tabla'] #Estas son las columnas con las que nos quedaremos

educacion['tabla'] = 'educacion' # Creamos la columna tabla como identificador
# Renombramos todas las columnas para tenerlas normalizadas
educacion = educacion.rename(columns={'Género': 'genero',
                                     'Función': 'cargo',
                                     'Total Ingreso': 'sueldo_bruto',
                                     'Sueldo Neto': 'sueldo_neto'})

educacion = educacion[columnas] #Filtramos las columnas que vamos a usar

salud['tabla'] = 'salud'
salud = salud.rename(columns={'SEXO': 'genero',
                              'CARGO': 'cargo',
                              'SUELDO BRUTO': 'sueldo_bruto',
                              'SUELDO NETO': 'sueldo_neto'})

salud = salud[columnas]

bancos['tabla'] = 'bancos'
bancos = bancos.rename(columns={'Género': 'genero',
                                'Posición Actual': 'cargo',
                                'Sueldo Bruto ': 'sueldo_bruto',
                                'Sueldo Neto ': 'sueldo_neto'})

bancos = bancos[columnas]
```

Codificación

Paso 4

- **Concatenación de las columnas con las que trabajaremos, eliminación de valores nulos y filtración de data**

```
# Luego de tener todas las tablas con las mismas columnas, concatenamos  
df = pd.concat([educacion,salud,bancos])
```

```
# Eliminamos valores nulos  
df = df.dropna()
```

```
# Filtramos la data basandonos en el sueldo minimo de RD y en un limite maximo de 100mil pesos  
df = df[df['sueldo_neto'] > 19352.50]  
df = df[df['sueldo_neto'] < 100000]
```

Paso 5

- Iniciamos con el análisis de la data. Obtenemos estadísticas descriptivas sobre los datos contenidos en nuestro DataFrame como lo son el recuento de valores, el promedio, la desviación estándar, valores mínimo y máximo y cuartiles.

Análisis

```
# Estadísticas básicas  
df.describe()
```

	sueldo_bruto	sueldo_netto
count	69364.000000	69364.000000
mean	54691.008901	40681.503849
std	18313.029198	14051.993964
min	19500.000000	19352.680000
25%	40000.000000	28651.312500
50%	50833.550000	39025.375000
75%	69663.100000	50335.190000
max	213281.060000	99928.980000

Codificación

Paso 6

- Función para graficación

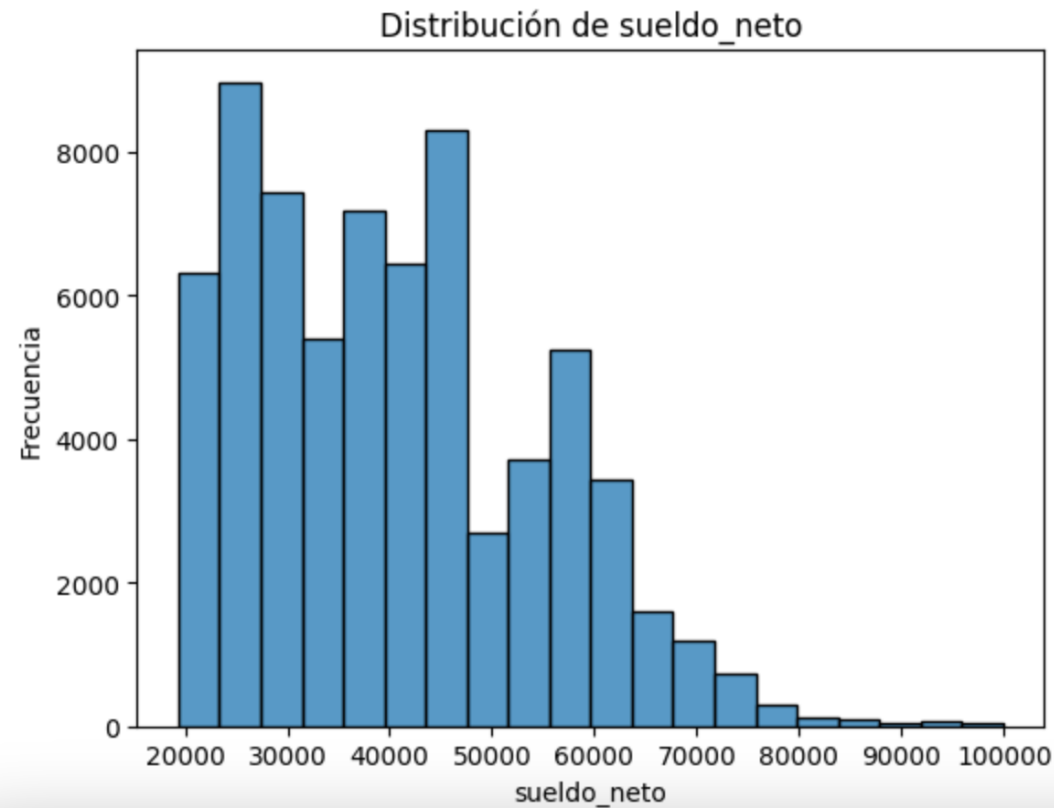
```
# Esta funcion recibe el nombre de una columna y la grafica en un histograma  
def freq_sueldos(columna):  
    sns.histplot(data=df, x=df[columna], bins=20) # Este metodo crea el histograma  
    plt.xlabel(columna) # Nombre del eje x  
    plt.ylabel('Frecuencia') # Nombre del eje y  
    plt.title(f'Distribución de {columna}') # Titulo general de la tabla  
    plt.show()
```

Codificación

Paso 7

- Histograma de la distribución del sueldo neto

```
freq_sueldos('sueldo_netos')
```

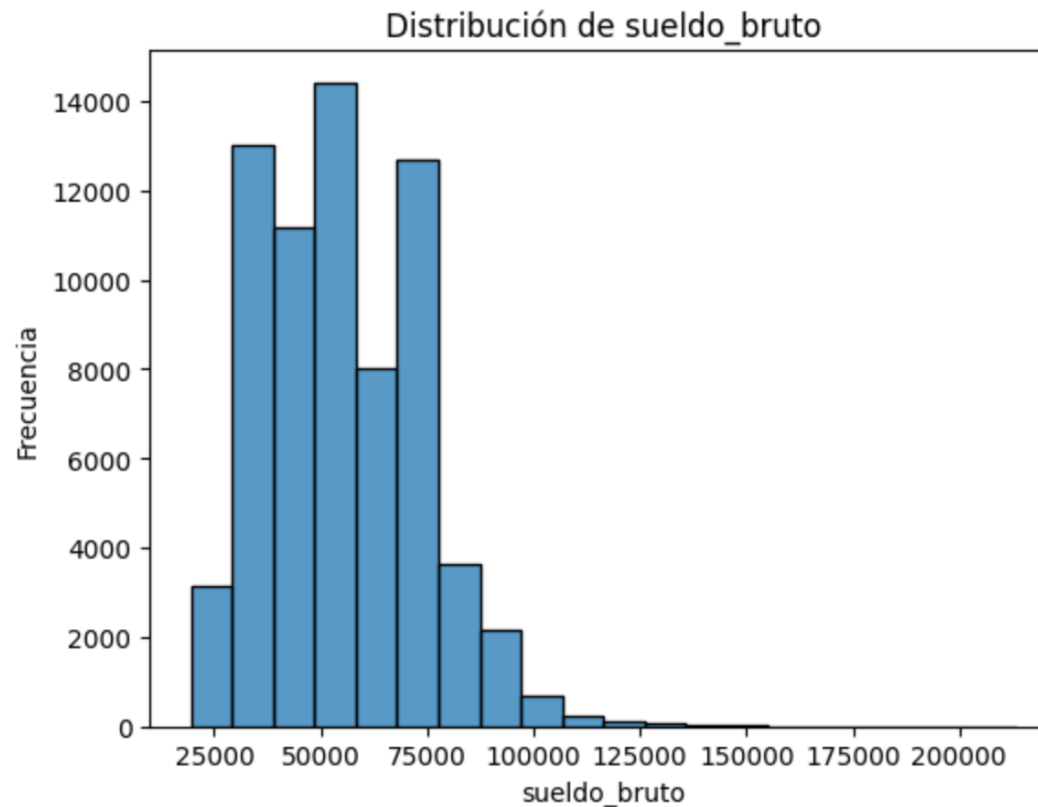


Codificación

Paso 8

- Histograma de la distribución del sueldo bruto

```
freq_sueldos('sueldo_bruto')
```

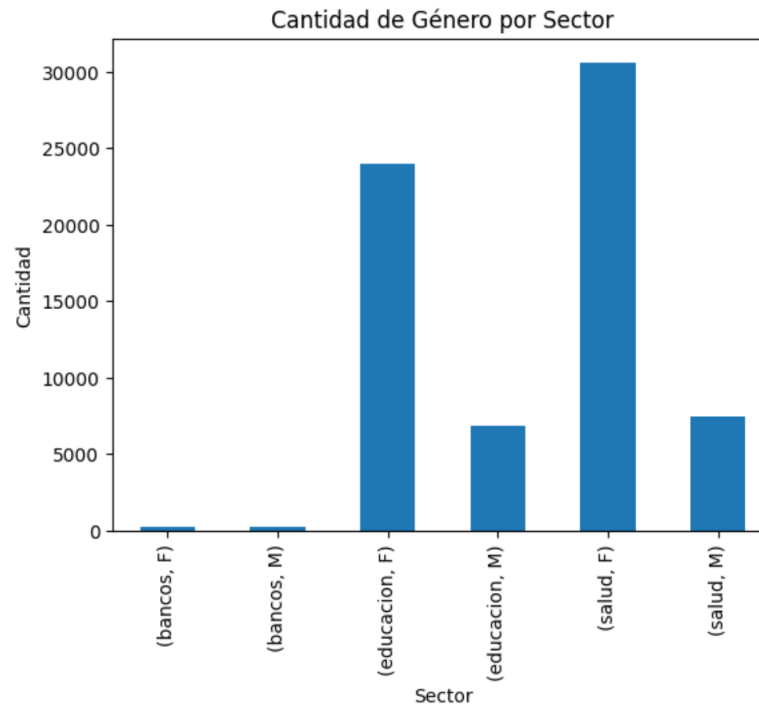


Codificación

Paso 9

- **Análisis de empleados por sector separados por género**

```
# Empleados por sector separados por genero
df_generos = df.groupby('tabla')['genero'].value_counts() # En una serie de Pandas guardamos la cantidad de persona por genero hay en cada sector
df_generos.plot(kind='bar') # Graficamos en diagrama de barras
plt.xlabel('Sector') # Nombramos los ejes
plt.ylabel('Cantidad')
plt.title('Cantidad de Género por Sector') # Título general
```



Codificación

Paso 10

- Promedio sueldo neto en cada sector

```
promedio_sn_educacion = df[df['tabla'] == 'educacion']['sueldo_net'].mean() # Tomamos el promedio del sueldo neto de cada sector
promedio_sn_bancos = df[df['tabla'] == 'bancos']['sueldo_net'].mean()
promedio_sn_salud = df[df['tabla'] == 'salud']['sueldo_net'].mean()

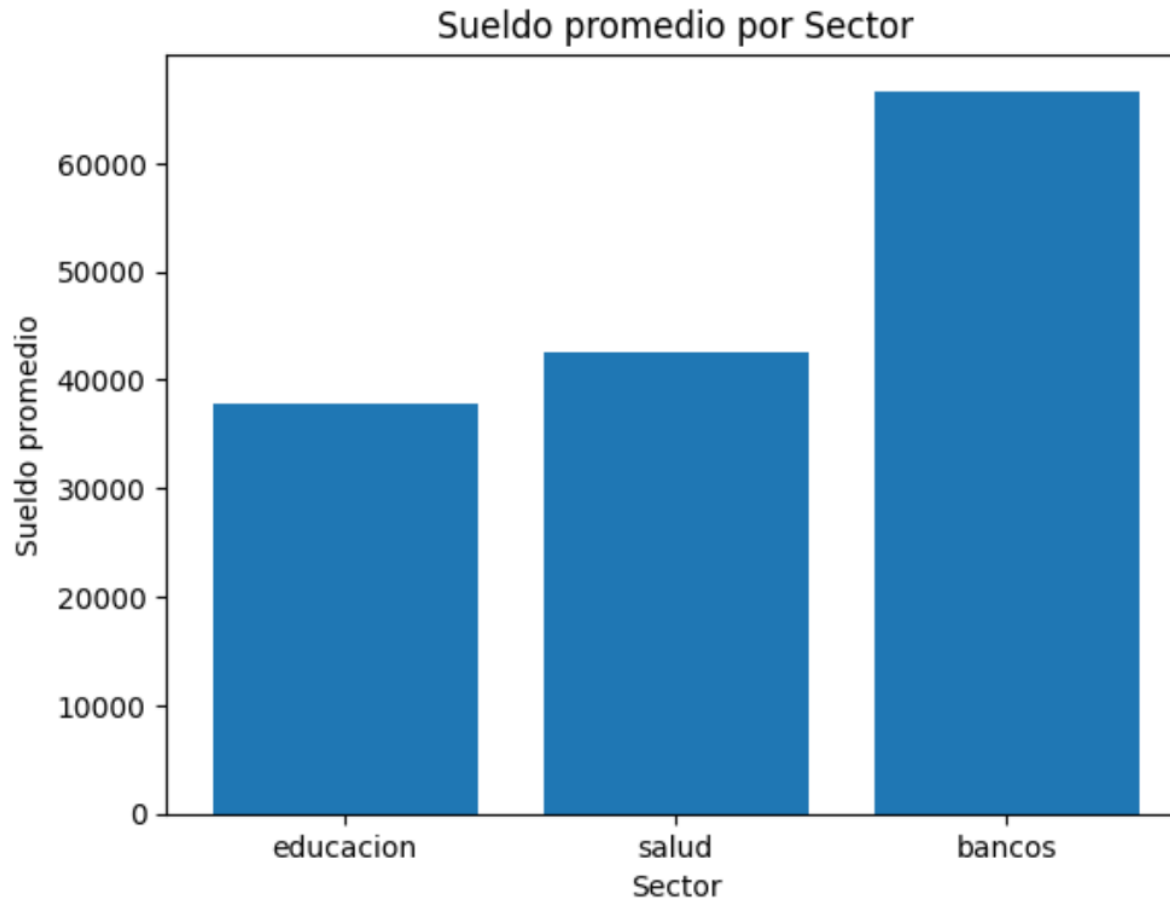
# Guardamos los promedios en una lista para graficarlos y en otra lista, siguiendo el orden, guardamos los nombres de los sectores para que :
promedios_sn_sector = [promedio_sn_educacion, promedio_sn_salud, promedio_sn_bancos]
sectores = ['educacion', 'salud', 'bancos']

plt.bar(sectores, promedios_sn_sector)
plt.xlabel('Sector') # Nombramos los ejes
plt.ylabel('Sueldo promedio')
plt.title('Sueldo promedio por Sector') # Titulo general
```

Codificación

Paso 10

- Gráfica promedio sueldo neto en cada sector



Codificación

Paso 11

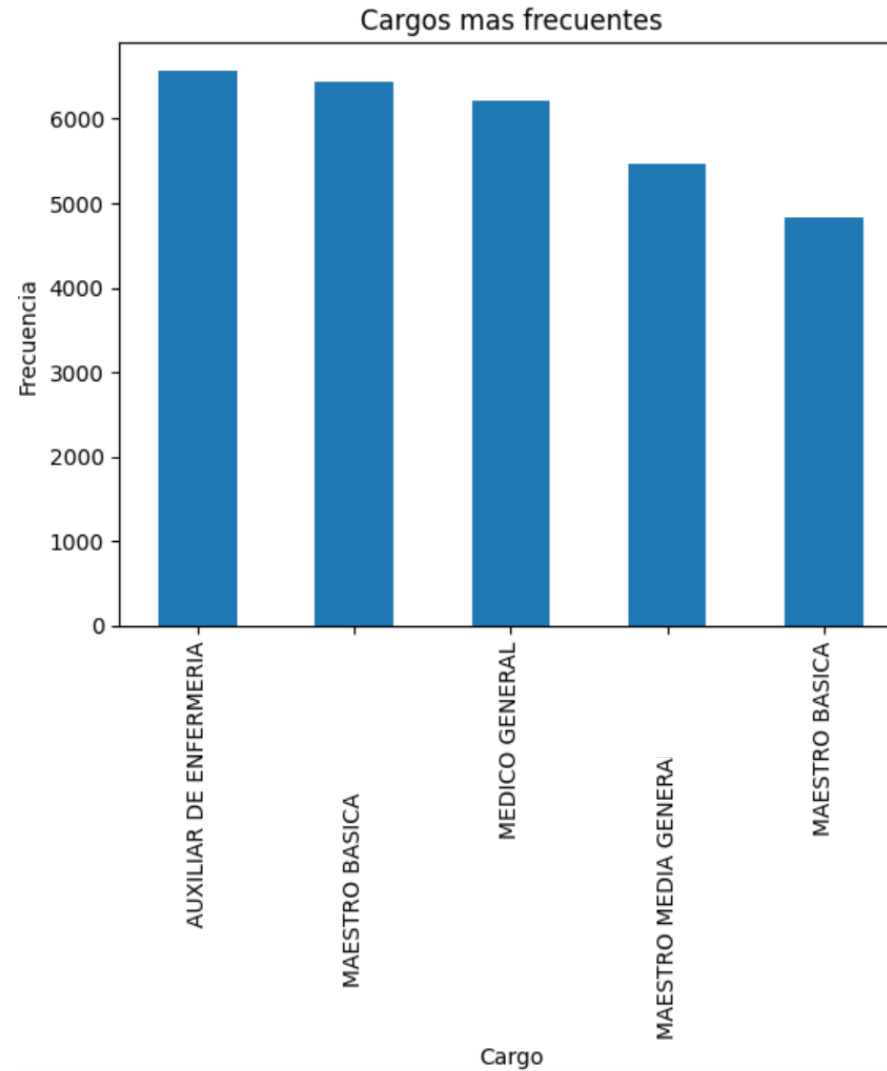
- Cargos más frecuentes

```
# 5 Cargos mas frecuentes
cargo_fq = df['cargo'].value_counts() # Contamos cuantas perdonas hay por cargo
cargo_fq = cargo_fq.head() # Sacamos un top 5
cargo_fq.plot.bar() # Graficamos con barras
plt.xlabel('Cargo') # Nombramos los ejes
plt.ylabel('Frecuencia')
plt.title('Cargos mas frecuentes') # Titulo general
```

Codificación

Paso 12

- Gráfica cargos más frecuentes





SAMSUNG

Together for Tomorrow!
Enabling People

Education for Future Generations

©2021 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.