

Examen parte 2

Eva Rebeca Mata Cavazos

ID 162375

Análisis Titanic

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [2]: # importar datos
titanic = pd.read_csv("titanic.csv")
titanic.head()
```

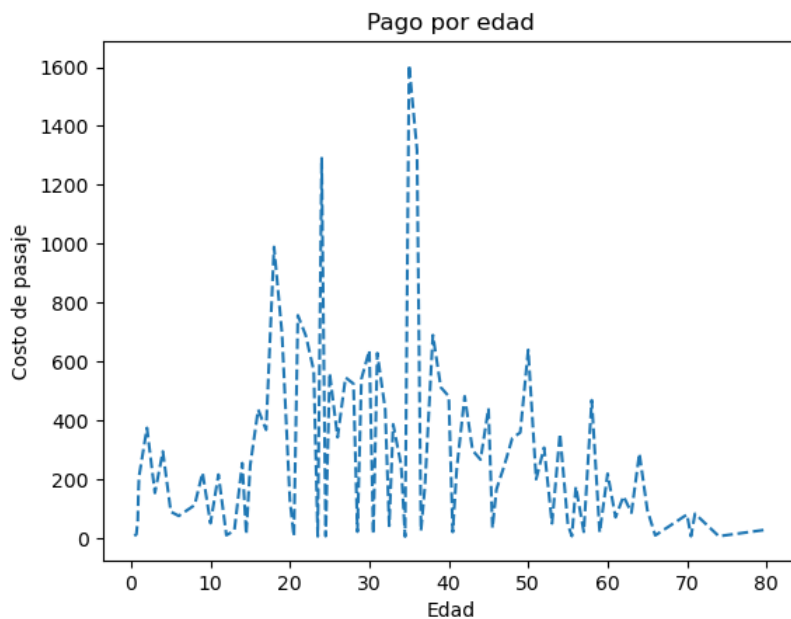
```
Out[2]:
```

	Unnamed: 0	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	3	1	3	Heikinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Gráfica de líneas

```
In [33]: # Comparando la edad contra lo que se pagó de pasaje. Se observa que entre los 30 y 40 se pagó más por el boleto.
# Curiosamente, hay muchos que no pagaron por abordar el titanic.

rel_fare_sur_df=titanic.groupby(["Age"]).sum().reset_index()
plt.plot(rel_fare_sur_df["Age"],rel_fare_sur_df["Fare"],linestyle="--")
plt.title("Pago por edad")
plt.xlabel("Edad")
plt.ylabel("Costo de pasaje")
plt.show()
```

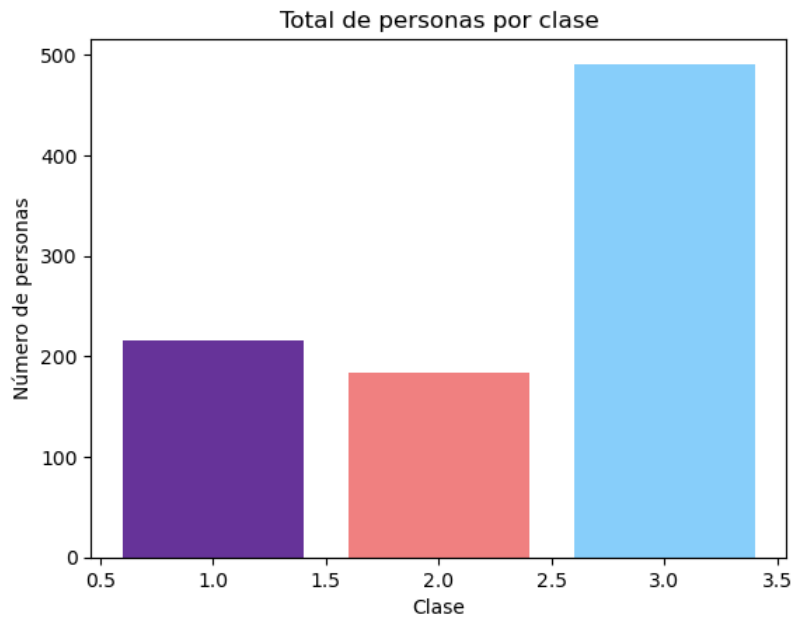


Gráfica de barras

```
In [35]: # Comparando las clases. Se podrá observar cuantos pertenecían en cada clase.

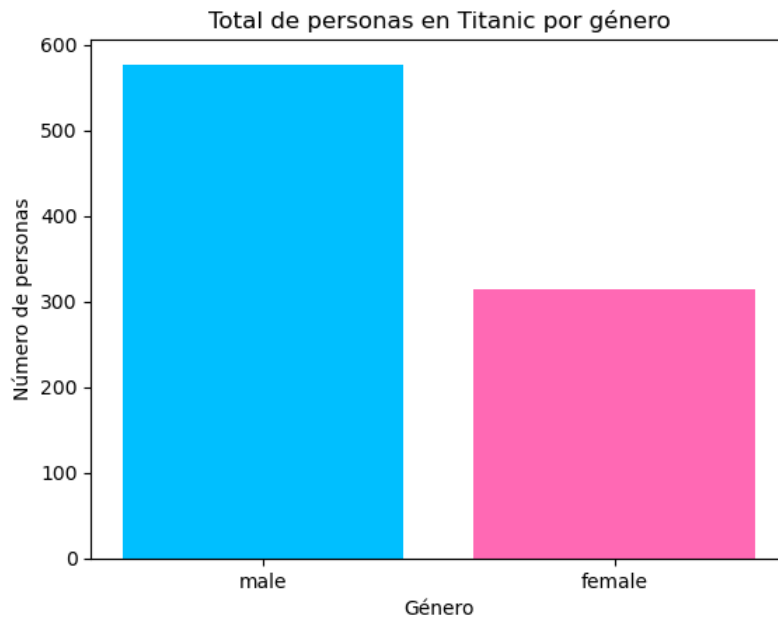
p_class=titanic.groupby("Pclass").count().reset_index()
```

```
p_class
plt.bar(p_class["Pclass"],p_class["Sex"],
        color=["rebeccapurple","lightcoral","lightskyblue"])
plt.title("Total de personas por clase")
plt.xlabel("Clase")
plt.ylabel("Número de personas")
plt.show()
```



In [36]: # Total de personas que había en el titanic entre hombres y mujeres.

```
cont_survived=titanic.groupby(["Sex"]).count().reset_index()
survived_sort=cont_survived.sort_values("Survived",ascending=False)
plt.bar(survived_sort["Sex"],survived_sort["Survived"],
        color=["deepskyblue","hotpink"])
plt.title("Total de personas en Titanic por género")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()
```



Gráfica tipo apilada (Stacked)

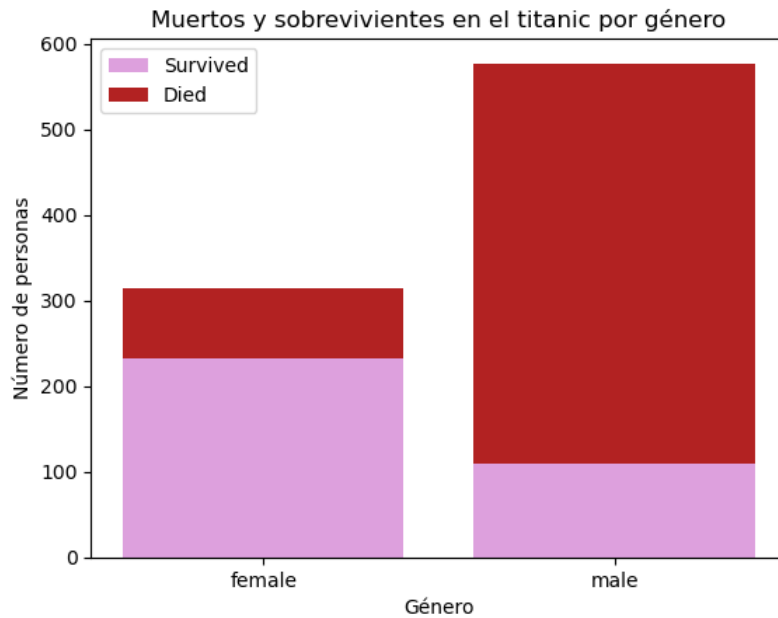
In [37]: # Total de sobrevivientes y muertos entre hombres y mujeres. Se puede ver que Las mujeres fueron Las que más se salvaron.

```
cont_sur_age=titanic.groupby(["Survived","Sex"]).count().reset_index()
cont_sur_age
cont_1=cont_sur_age[cont_sur_age['Survived']!=0]
cont_1.head() #sobrevivieron
```

```

cont_0=cont_sur_age[cont_sur_age['Survived']!=1]
cont_0.head() #murieron
plt.bar(cont_1["Sex"],cont_1["Pclass"],label="Survived",color="plum")
plt.bar(cont_0["Sex"],cont_0["Pclass"],label="Died",bottom=cont_1["Pclass"],color="firebrick")
plt.legend()
plt.title("Muertos y sobrevivientes en el titanic por género")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()

```



Gráfica tipo grouped

In [38]: *##Muestra mejor la cantidad de hombres y mujeres que sobrevivieron y murieron en el titanic.
#La diferencia en hombres es bastante.*

```

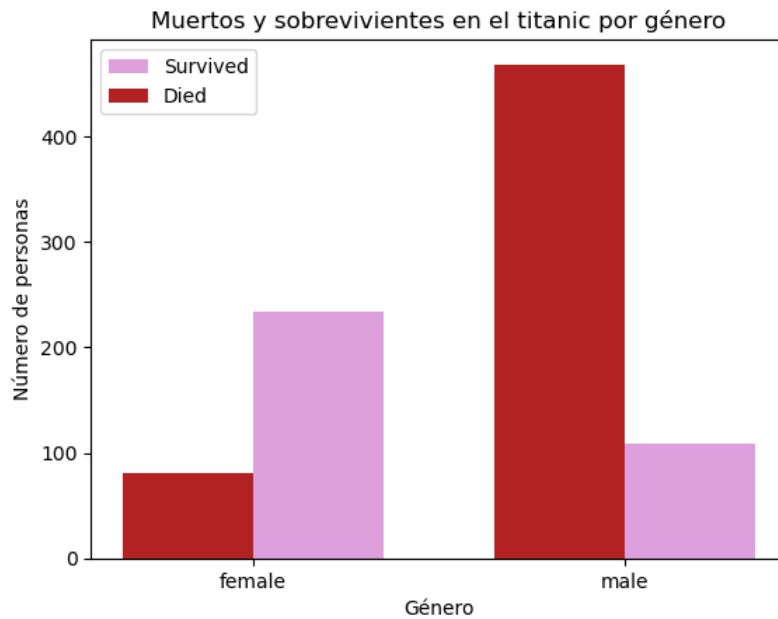
width=0.35 # ancho de barras
labels=cont_1["Sex"]
x=np.arange(len(labels))

#pintando las barras
plt.bar(x+width/2,cont_1["Pclass"],width,label="Survived",color="plum")
plt.bar(x-width/2,cont_0["Pclass"],width,label="Died",color="firebrick")

# cambiando el eje x a los países
plt.xticks(x,labels)

#incluyendo leyend
plt.legend()
plt.title("Muertos y sobrevivientes en el titanic por género")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()

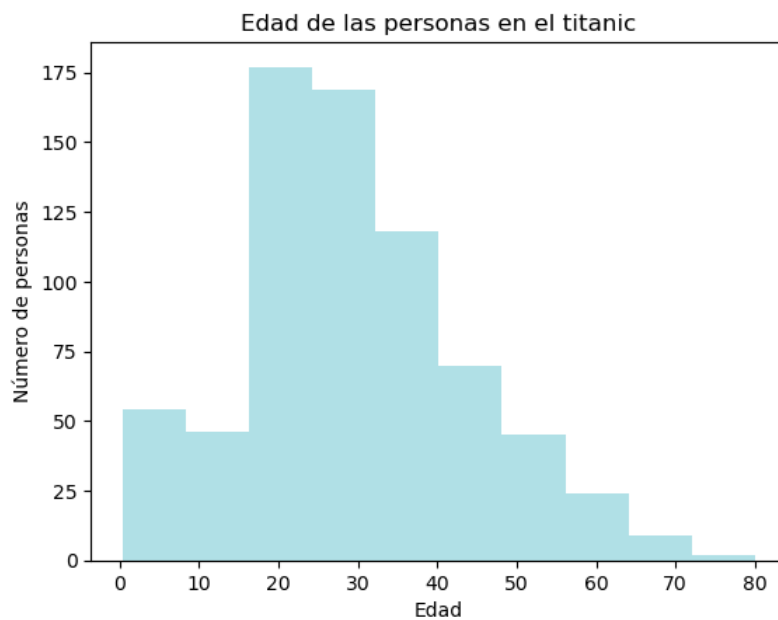
```



Histograma básico

In [39]: *# La mayor parte de personas que estaban en el titanic tenían entre 20 y 30 años.*

```
plt.hist(titanic["Age"],color="powderblue")
plt.title("Edad de las personas en el titanic")
plt.xlabel("Edad")
plt.ylabel("Número de personas")
plt.show()
```



Histograma agrupado

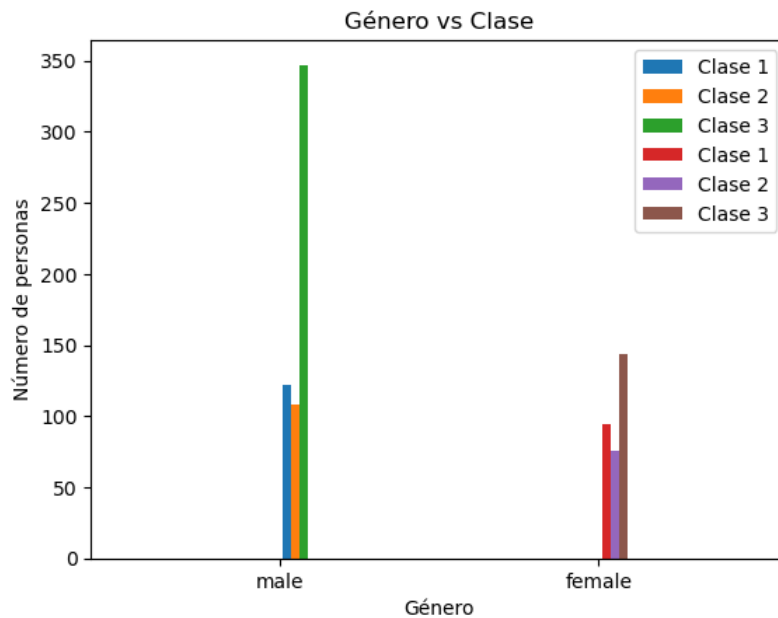
In [31]: *# Clase comprada por sexo. Se observa que mucho hombres compraron en 3ra clase.*

```
fem_class1=titanic[(titanic['Sex']=="female") & (titanic['Pclass']==1)]
fem_class2=titanic[(titanic['Sex']=="female") & (titanic['Pclass']==2)]
fem_class3=titanic[(titanic['Sex']=="female") & (titanic['Pclass']==3)]

male_class1=titanic[(titanic['Sex']=="male") & (titanic['Pclass']==1)]
male_class2=titanic[(titanic['Sex']=="male") & (titanic['Pclass']==2)]
male_class3=titanic[(titanic['Sex']=="male") & (titanic['Pclass']==3)]

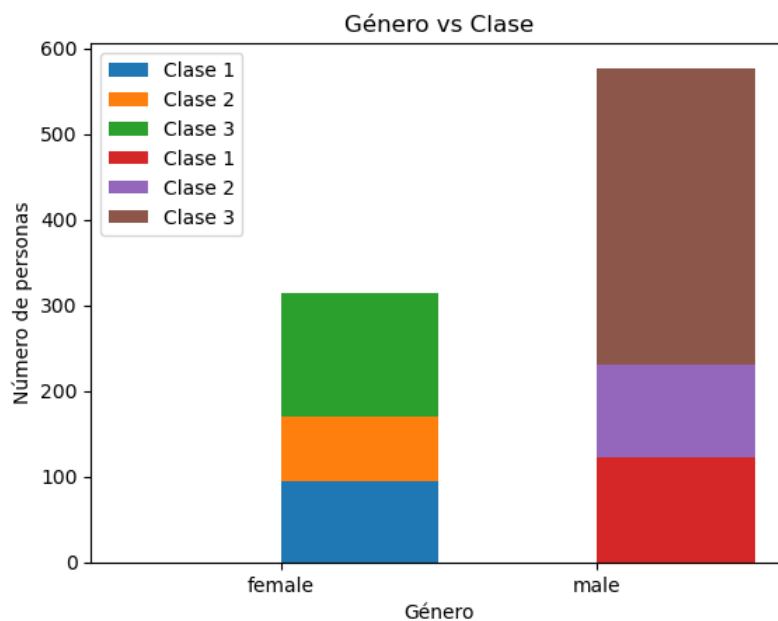
plt.hist([male_class1["Sex"],male_class2["Sex"],male_class3["Sex"]],
         label=["Clase 1","Clase 2","Clase 3"])
plt.hist([fem_class1["Sex"],fem_class2["Sex"],fem_class3["Sex"]],
```

```
label=["Clase 1","Clase 2","Clase 3"])
plt.legend()
plt.title("Género vs Clase")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()
```



```
In [40]: # Al sobreponerse las gráficas se ve una mayor cantidad de hombres de tercera clase que mujeres.
# Así como una mayor cantidad de hombres de primera clase que mujeres.

plt.hist([fem_class1["Sex"],fem_class2["Sex"],fem_class3["Sex"]],
         label=["Clase 1","Clase 2","Clase 3"],
         width=0.5,stacked=True)
plt.hist([male_class1["Sex"],male_class2["Sex"],male_class3["Sex"]],label=["Clase 1","Clase 2","Clase 3"],
         width=0.5,stacked=True)
plt.legend()
plt.title("Género vs Clase")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()
```

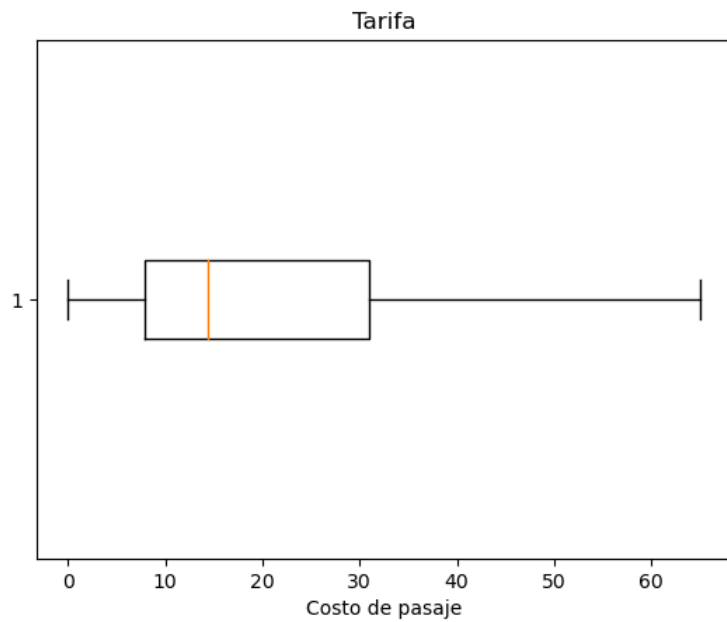


Boxplot

```
In [42]: # La mediana mide La tendencia central, en este caso La tarifa de todos los que estaban en el titanic.

plt.boxplot(titanic["Fare"],showfliers=False,
            vert=False)
```

```
plt.title("Tarifa")
plt.xlabel("Costo de pasaje")
plt.show()
```

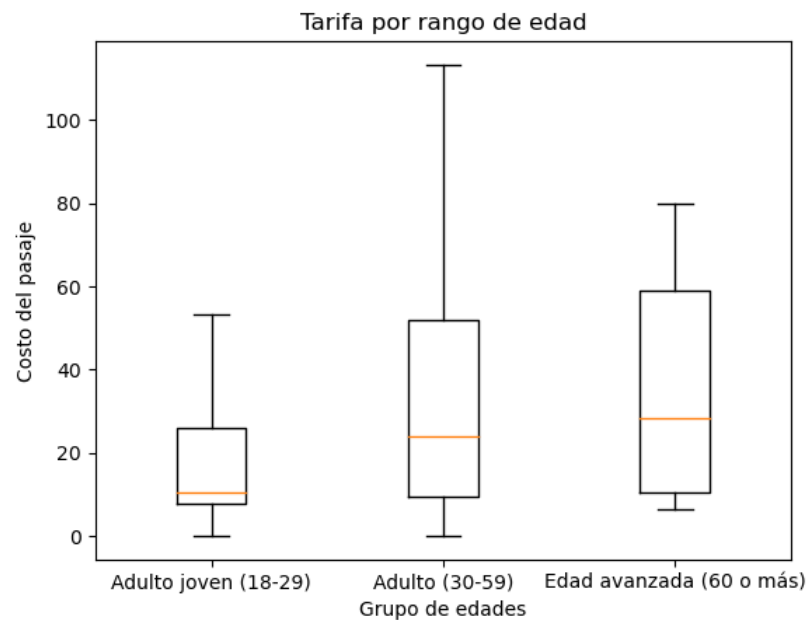


Boxplot agrupado

In [43]: *# Tarifa por rango de edad, Los de tercera edad fueron los que pagaron más en el titanic.*

```
adulto_joven_df=titanic[(titanic["Age"]>=18) & (titanic["Age"]<30)]
adulto_df=titanic[(titanic["Age"]>=30) & (titanic["Age"]<60)]
adulto_avanzado_df=titanic[titanic["Age"]>=60]

plt.boxplot([adulto_joven_df["Fare"],adulto_df["Fare"],adulto_avanzado_df["Fare"]],
            labels=["Adulto joven (18-29)","Adulto (30-59)","Edad avanzada (60 o más)"],
            showfliers=False)
plt.title("Tarifa por rango de edad")
plt.xlabel("Grupo de edades")
plt.ylabel("Costo del pasaje")
plt.show()
```



Análisis Lungcap

In [13]: *# importar datos*

```
lungcap = pd.read_csv("lungcap.csv")
lungcap.head()
```

```
Out[13]:
```

	Unnamed: 0	Age	FEV	Ht	Gender	Smoke
0	1	3	1.072	46.0	F	0
1	2	4	0.839	48.0	F	0
2	3	4	1.102	48.0	F	0
3	4	4	1.389	48.0	F	0
4	5	4	1.577	49.0	F	0

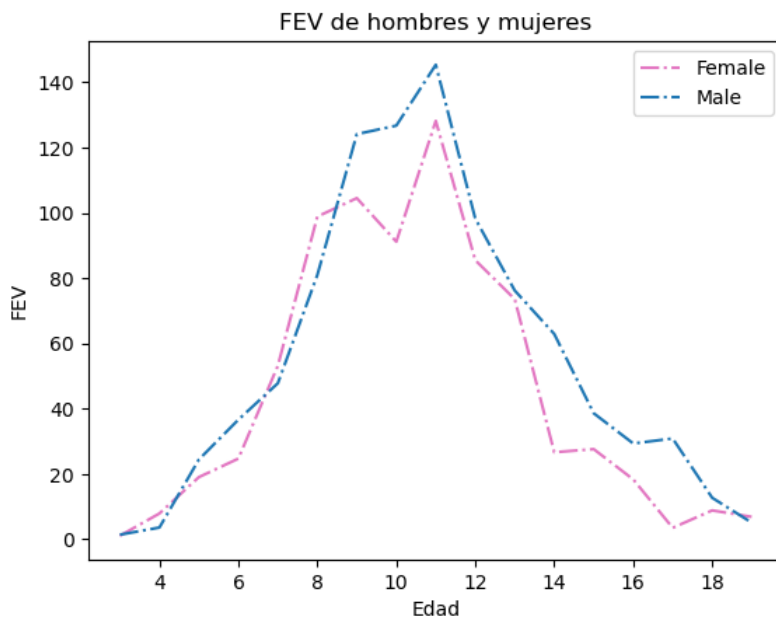
Gráfica de líneas

```
In [44]: # FEV por edad de hombres y mujeres. Ambos sexos tienen un mayor FEV entre Los 10 y 12 años.

suma_edad_df=lungcap.groupby(["Age", "Gender"]).sum().reset_index()
suma_edad_df

suma_edad_f=suma_edad_df[suma_edad_df["Gender"]=="F"]
suma_edad_m=suma_edad_df[suma_edad_df["Gender"]=="M"]

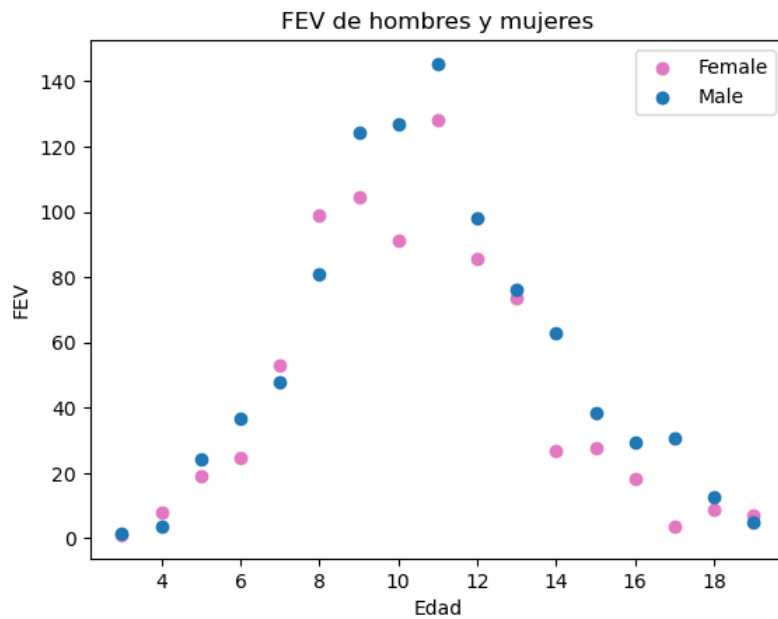
plt.plot(suma_edad_f["Age"], suma_edad_f["FEV"], label="Female", color="tab:pink", linestyle="-.-")
plt.plot(suma_edad_f["Age"], suma_edad_m["FEV"], label="Male", color="#1f77b4", linestyle="-.-")
plt.legend()
plt.title("FEV de hombres y mujeres")
plt.xlabel("Edad")
plt.ylabel("FEV")
plt.show()
```



Scatter plot

```
In [45]: # Otra forma de observar La data anterior.

plt.scatter(suma_edad_f["Age"], suma_edad_f["FEV"], label="Female", color="tab:pink")
plt.scatter(suma_edad_m["Age"], suma_edad_m["FEV"], label="Male", color="#1f77b4")
plt.legend()
plt.title("FEV de hombres y mujeres")
plt.xlabel("Edad")
plt.ylabel("FEV")
plt.show()
```

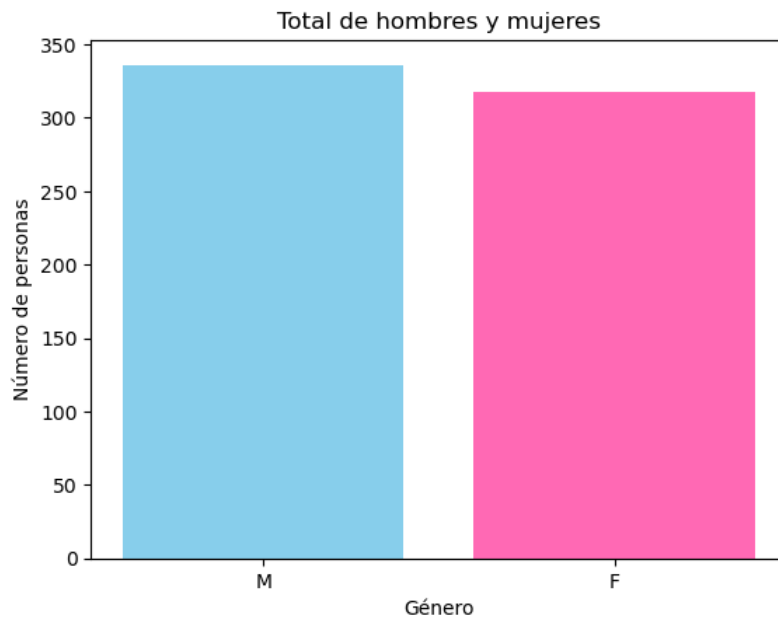


Gráfica de barras

In [46]: *# Total de hombres y mujeres en la base de datos. Hay un poco más de hombres en este estudio.*

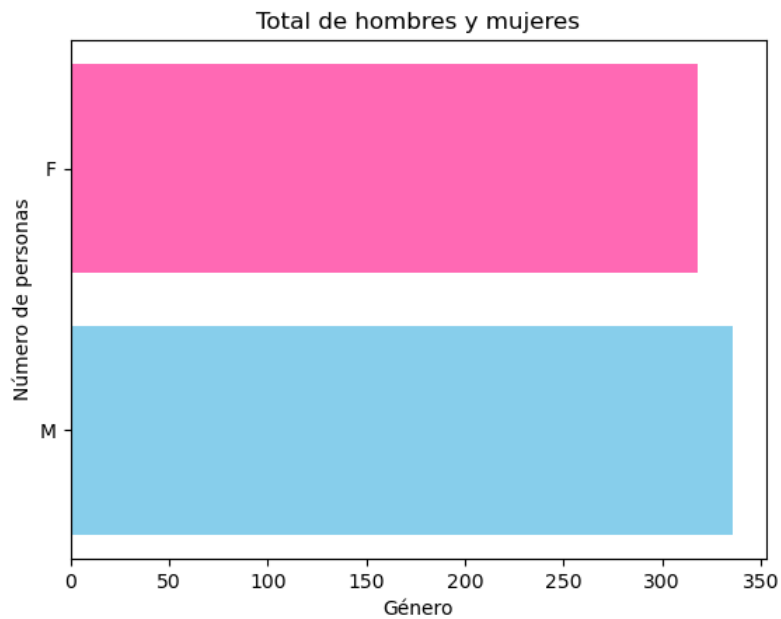
```
total_male_fem=lungcap.groupby(["Gender"]).count().reset_index()
total_sort=total_male_fem.sort_values("Smoke",ascending=False)

plt.bar(total_sort["Gender"],total_sort["Smoke"],
        color=["skyblue","hotpink"])
plt.title("Total de hombres y mujeres")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()
```



In [47]: *# Otra forma de ver la información anteriormente presentada.*

```
plt.barh(total_sort["Gender"],total_sort["Smoke"],
        color=["skyblue","hotpink"])
plt.title("Total de hombres y mujeres")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()
```

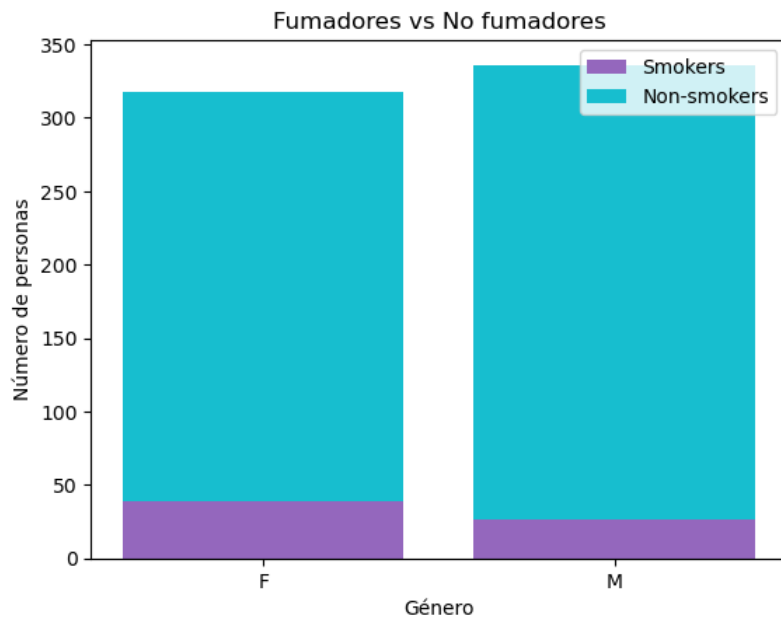
Gráfica tipo apilada (Stacked)

```
In [50]: # Muestra la cantidad de fumadores y no fumadores por hombres y mujeres.

cont_smo_gen=lungcap.groupby(["Smoke","Gender"]).count().reset_index()
cont_smo_gen

cont_1_smo=cont_smo_gen[cont_smo_gen['Smoke']!=0]
cont_1_smo.head() #fuman
cont_0_smo=cont_smo_gen[cont_smo_gen['Smoke']!=1]
cont_0_smo.head() #no fuman

plt.bar(cont_1_smo["Gender"],cont_1_smo["Age"],label="Smokers",color="tab:purple")
plt.bar(cont_0_smo["Gender"],cont_0_smo["Age"],label="Non-smokers",bottom=cont_1_smo["Age"],color="tab:cyan")
plt.legend()
plt.title("Fumadores vs No fumadores")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()
```



Gráfica tipo grouped

```
In [49]: # Otra forma de expresar la información de arriba.

width=0.35 # ancho de barras
```

```

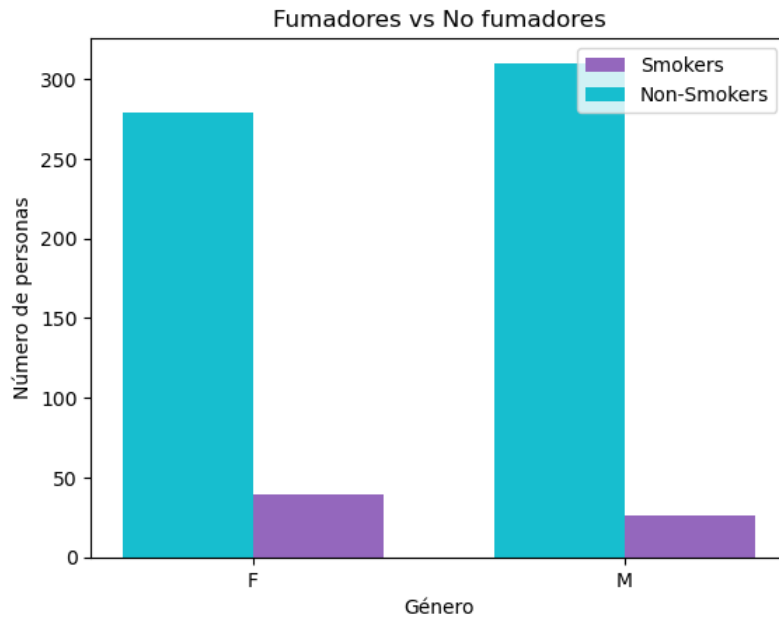
labels=cont_1_smo["Gender"]
x=np.arange(len(labels))

#pintando las barras
plt.bar(x+width/2,cont_1_smo["Age"],width,label="Smokers",color="tab:purple")
plt.bar(x-width/2,cont_0_smo["Age"],width,label="Non-Smokers",color="tab:cyan")

# cambiando el eje x a los paises
plt.xticks(x,labels)

#incluyendo leyend
plt.legend()
plt.title("Fumadores vs No fumadores")
plt.xlabel("Género")
plt.ylabel("Número de personas")
plt.show()

```



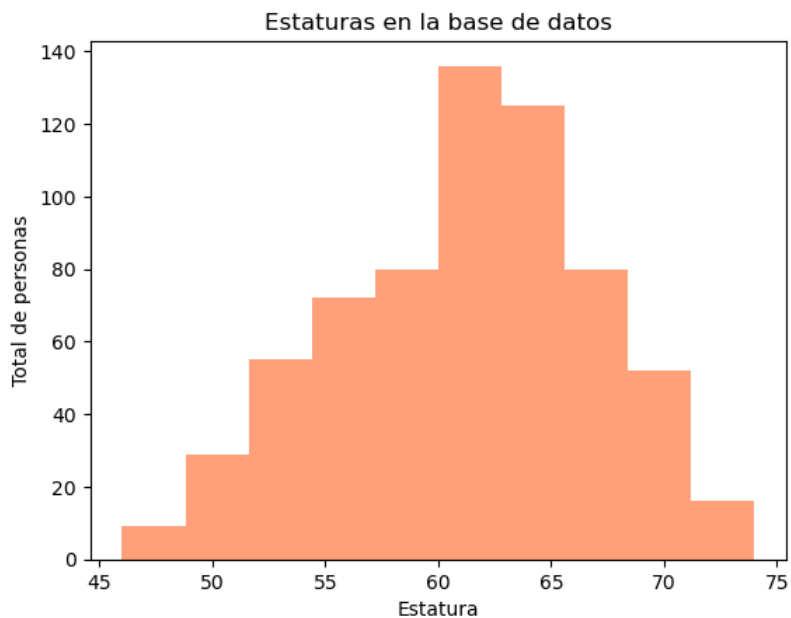
Histograma

```

In [51]: # La altura que tienen las personas en la base de datos.
# La mayor parte se encuentra entre los 60 y 64 inches.

plt.hist(lungcap["Ht"],color="lightsalmon")
plt.title("Estaturas en la base de datos")
plt.xlabel("Estatura")
plt.ylabel("Total de personas")
plt.show()

```

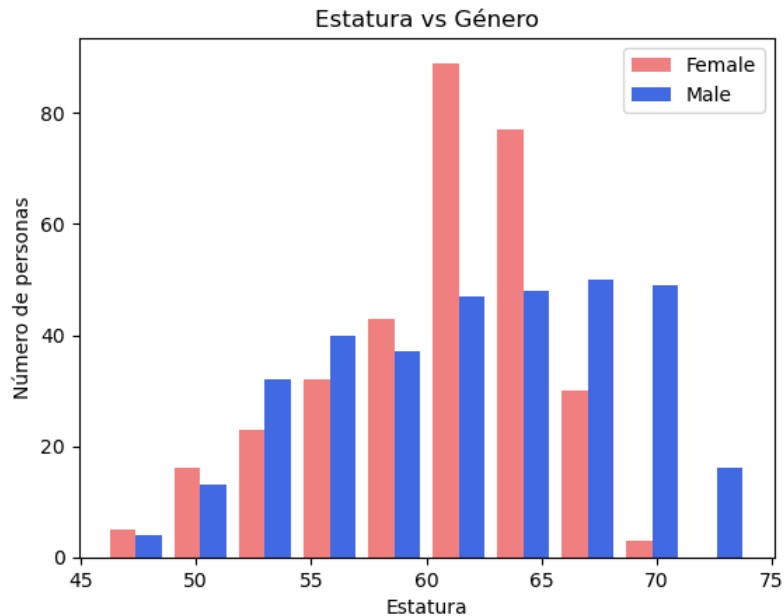


Histograma agrupado

```
In [52]: # Comparando La estatura con el género. Son más altas Las mujeres que Los hombres.

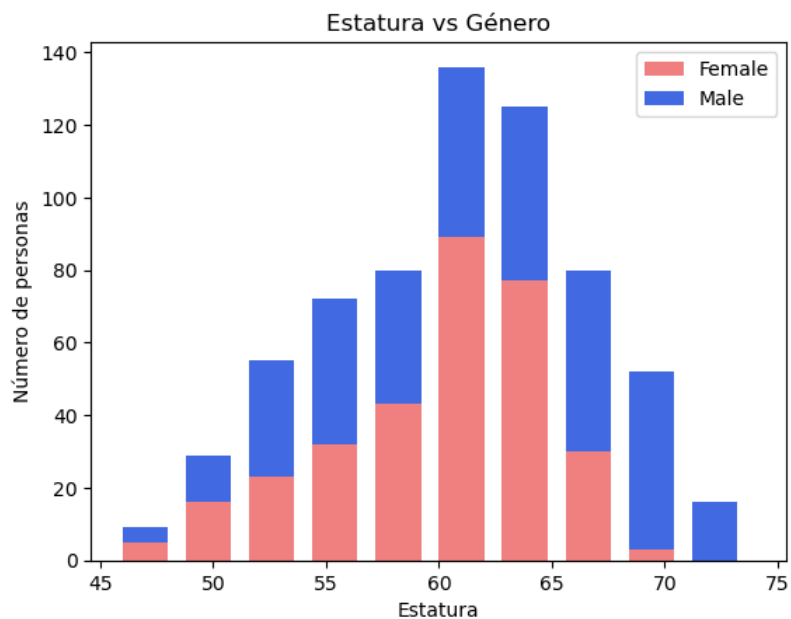
lung_f=lungcap[lungcap["Gender"]=="F"]
lung_m=lungcap[lungcap["Gender"]=="M"]

plt.hist([lung_f["Ht"],lung_m["Ht"]],label=["Female","Male"],
         color=["lightcoral","royalblue"])
plt.legend()
plt.title("Estatura vs Género")
plt.xlabel("Estatura")
plt.ylabel("Número de personas")
plt.show()
```



```
In [53]: # Otra forma de expresar La estatura y el género en conjunto.
```

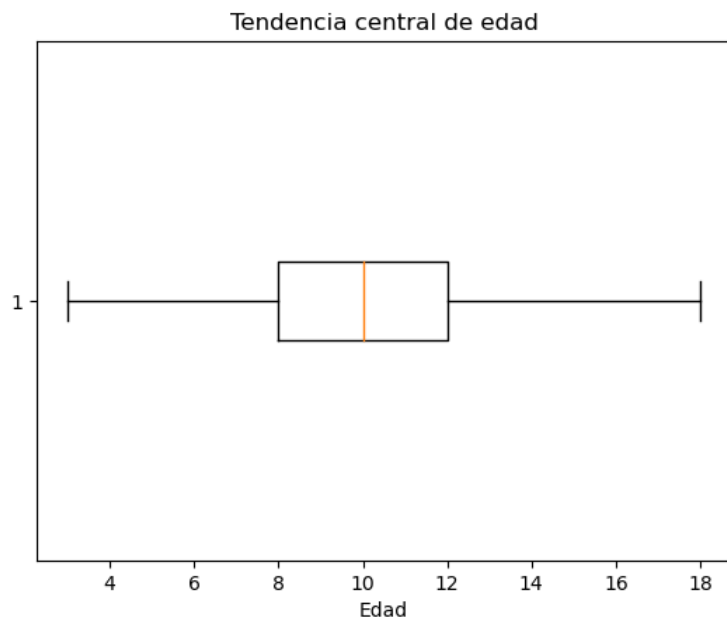
```
plt.hist([lung_f["Ht"],lung_m["Ht"]],label=["Female","Male"],
         width=2,stacked=True,color=["lightcoral","royalblue"])
plt.legend()
plt.title("Estatura vs Género")
plt.xlabel("Estatura")
plt.ylabel("Número de personas")
plt.show()
```



Boxplot

In [56]: *#Muestra Las edades del total en La base de datos. La tendencia central está entre los 8 y 12 años.*

```
plt.boxplot(lungcap["Age"],showfliers=False,
            vert=False)
plt.title("Tendencia central de edad")
plt.xlabel("Edad")
plt.show()
```



Boxplot agrupado

In [57]: *#Tendencia central de La edad de Los que fuman y no fuman. Se puede observar*

```
smokers=lungcap.groupby(["Smoke","Age"]).count().reset_index()
smokers

smokers_1=smokers[smokers['Smoke']!=0]
smokers_1.head() #fuman
non_smokers_0=smokers[smokers['Smoke']==1]
non_smokers_0.head()

plt.boxplot([smokers_1["Age"],non_smokers_0["Age"]],
            labels=["Fumadores","No fumadores"])
plt.title("Fumadores vs No fumadores")
plt.ylabel("Edades")
plt.show()
```

