

```
[In ]9: import csv
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv("C:\Users\vnacha\Desktop\Coding_Study\Portfolio\Insurance.csv")

[In ]18: sns.displot(df)

age    sex    bmi    children    smoker    region    charges
0      19  female  27.900      0      yes  southwest  16884.9240
1      18  male   31.770      1      no   southeast  3725.5520
2      28  male   31.000      3      no   southeast  4449.4620
3      33  male   22.705      0      no   northwest  21984.4761
4      32  male   28.880      0      no   northwest  3964.8520
...    ...    ...    ...    ...    ...    ...
1333   50  male   30.970      3      no   northwest  15600.5430
1334   18  female  31.920      0      no   northeast  2205.9080
1335   18  female  36.850      0      no   southeast  1629.8330
1336   21  female  25.800      0      no   southwest  2007.9450
1337   61  female  29.070      0      yes  northeast  2911.3600
1338 rows x 7 columns

[In ]29: sns.set_style("white")
sns.set_palette("pastel")
sns.relplot(x="age",
            y="charges",
            data=df,
            col="smoker",
            row="region",
            facet_kws={"col_wrap": 2})

[In ]23: sex = df["sex"]
sns.displot(sex)
plt.show()

[In ]24: children = df["children"]
sns.displot(children)
plt.show()

[In ]25: x = df["age"]
y = df["charges"]
plt.scatter(x, y, alpha=0.5)
sns.regplot(data=df, x="age", y="charges")
plt.show()

[In ]26: region = df["region"]
sns.displot(region)
plt.show()

[In ]27: smoker = df["smoker"]
sns.displot(smoker)
plt.show()

[In ]38: sns.violinplot(data=df, x="region", y="charges")
plt.show()

[In ]32: region_mean_age = df.groupby("region").age.mean()
region_mean_age.head()

[In ]36: region_mean_charges = df.groupby("region").charges.mean()
region_mean_charges.head()

[In ]37: sns.violinplot(data=df, x="region", y="children")
plt.show()

[In ]54: children_40 = []
for i in df["age"]:
    if i == 40:
        children_40.append(df["children"])
print(len(children_40))

[In ]60: new_df = df.loc[df["age"] == 40]
print(new_df)

[In ]66: children_40 = []
children_40.extend(new_df["children"].tolist())
print(children_40)

[In ]68: ave_children = sum(children_40)/len(children_40)
print(ave_children)

[In ]76: bmi_list = []
bmi_list.extend(df["bmi"].tolist())
ave_bmi = sum(bmi_list)/len(bmi_list)
print(ave_bmi)

[In ]84: no_of_people_above_ave = 0
for i in bmi_list:
    if i > ave_bmi:
        no_of_people_above_ave += 1
print(no_of_people_above_ave)
print(round(no_of_people_above_ave/len(bmi_list)*100,2))

[In ]135: age = []
sex = []
children = []
smoker = []
region = []
charges = []

age.extend(df["age"].tolist())
sex.extend(df["sex"].tolist())
children.extend(df["children"].tolist())
smoker.extend(df["smoker"].tolist())
region.extend(df["region"].tolist())
charges.extend(df["charges"].tolist())

combined_list = list(zip(age, sex, bmi_list, children, smoker, region, charges))
#print(combined_list)

[In ]153: Dict_Con = {}
for i in range(0, len(combined_list)):
    Dict_Con["Customer {}".format(i)] = combined_list[i]
#print(Dict_Con)

[In ]154: print(Dict_Con["Customer 0"])

[In ]94: print(Dict_Con["Customer 36"])

[In ]95: print(Dict_Con["Customer 44"])

[In ]96: x = df["age"]
y = df["charges"]
plt.scatter(x, y, alpha=0.5)
sns.regplot(data=df, x="age", y="charges")
plt.show()

[In ]98: x = df["bmi"]
y = df["charges"]
plt.scatter(x, y, alpha=0.5)
sns.regplot(data=df, x="bmi", y="charges")
plt.show()

[In ]128: bmi_charge_list = list(zip(bmi_list, charges))
bmi_charge_list_above_ave = []
for i in bmi_charge_list:
    if i[0] > ave_bmi:
        bmi_charge_list_above_ave.append(i[1])
print(bmi_charge_list_above_ave)
print(sum(bmi_charge_list_above_ave)/len(bmi_charge_list_above_ave))
print(len(bmi_charge_list_above_ave))

[In ]132: x = bmi_large_than_ave_charge
y = charge_for_bmi_above_ave
plt.scatter(x, y, alpha=0.5)
sns.regplot(x=bmi, y=charges)
plt.show()

[In ]124: new_df_1 = df.loc[df["charges"] > 30000]
print(new_df_1)
print(len(new_df_1))

[In ]125: smoker = new_df_1["smoker"]
sns.displot(smoker)
plt.show()

[In ]127: number_of_smoker = 0
for i in new_df_1["smoker"]:
    if i == "yes":
        number_of_smoker += 1
print(number_of_smoker)

[In ]129: perc_number_of_smoker = (number_of_smoker/len(new_df_1))*100
print(perc_number_of_smoker)

[In ]130: number_high_bmi = 0
for i in new_df_1["bmi"]:
    if i > ave_bmi:
        number_high_bmi += 1
print(number_high_bmi)
perc_number_high_bmi = (number_high_bmi/len(new_df_1))*100
print(perc_number_high_bmi)

[In ]136: sex = new_df_1["sex"]
sns.displot(sex)
plt.show()

[In ]134: region = new_df_1["region"]
sns.displot(region)
plt.show()

[In ]140: print(len(new_df_1["children"])/len(new_df_1["children"]))
print(len(new_df_1["children"]))
print(sum(df["children"])/len(df["children"]))
print(len(df["children"]))
print(len(new_df_1["children"]))

[In ]141: age_1 = new_df_1["age"]
sns.displot(age_1)
sns.despine()
plt.show()

[In ]145: smoker_list = []
sex_list = []
smoker_list.extend(new_df_1["smoker"].tolist())
sex_list.extend(new_df_1["sex"].tolist())

count_smoker = list(zip(smoker_list, sex_list))
#print(count_smoker)

count = 0
for i in count_smoker:
    if i[0] == "yes" and i[1] == "male":
        count += 1
print(count)

count = 0
for i in count_smoker:
    if i[0] == "yes" and i[1] == "female":
        count += 1
print(count)

[In ]148: print(count/number_of_smoker)

[In ]147: count_female = 0
for i in count_smoker:
    if i[0] == "yes" and i[1] == "female":
        count_female += 1
print(count_female)

[In ]149: print(count/number_of_smoker)

[In ]150: male = 0
for i in sex_list:
    if i == "male":
        male += 1
print(male)

[In ]151: female = 0
for i in sex_list:
    if i == "female":
        female += 1
print(female)

[In ]152: new_df_2 = new_df_1.loc[df["smoker"] == "no"]
print(new_df_2)
print(len(new_df_2))

[In ]1: 
```