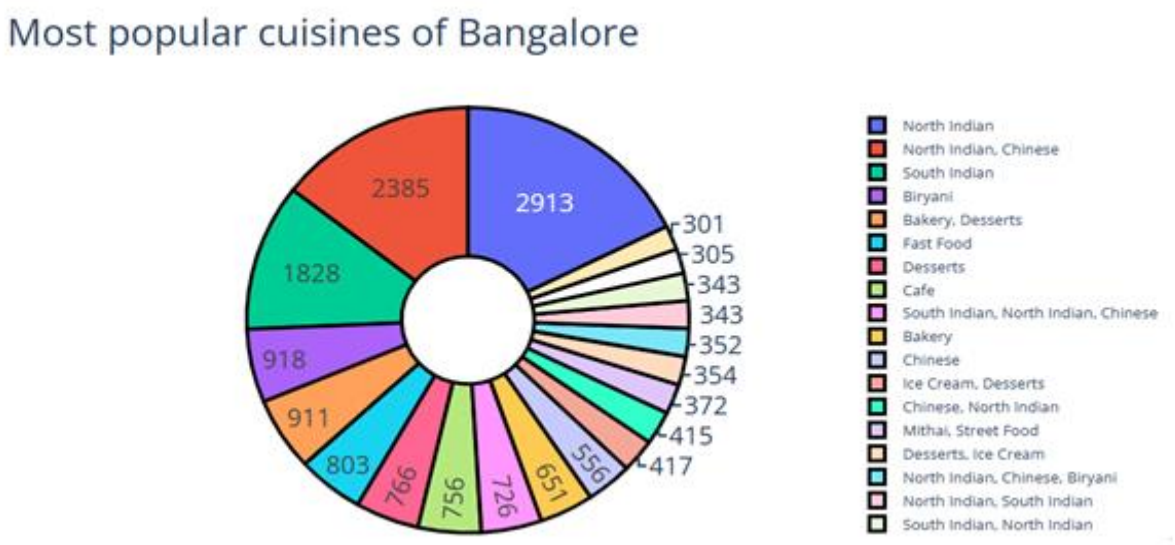
Mastering Data Visualization Techniques (Part 5)

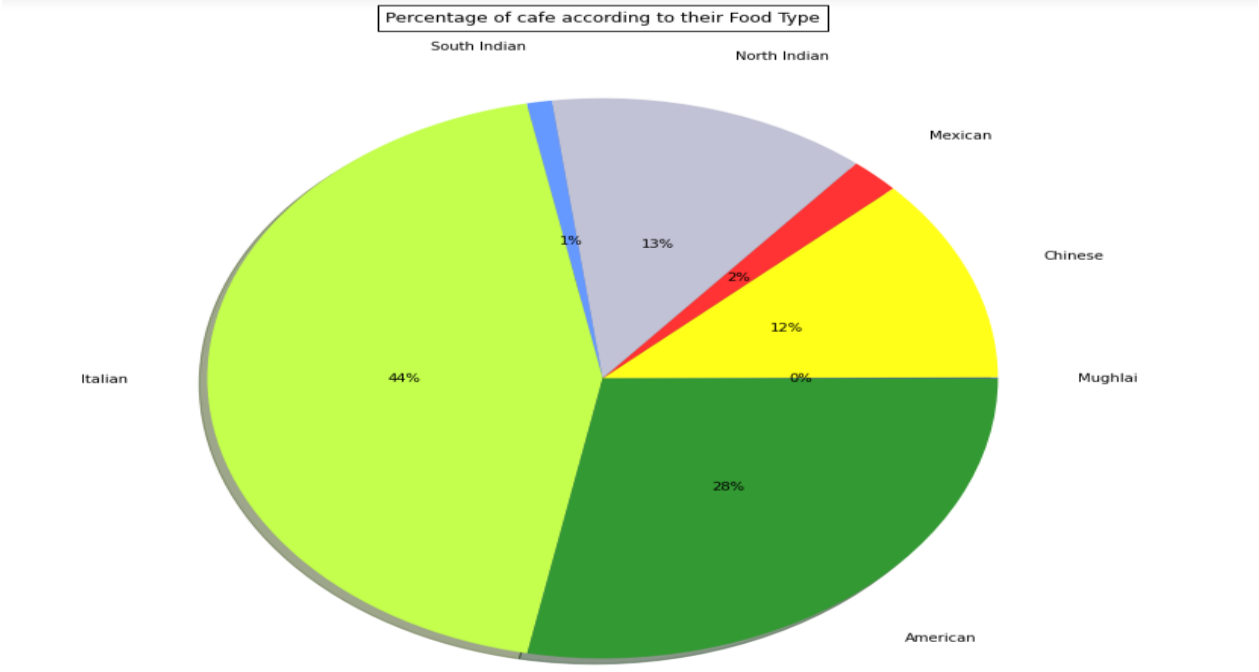
Prepared by: Syed Afroz Ali

```
import plotly.graph_objs as go
values = data['cuisines'].value_counts()[:20]
labels=values.index
text=values.index
fig =
go.Figure(data=[go.Pie(values=values,labels=labels,hole=.3)]
)
fig.update_traces(hoverinfo='label+percent', textinfo='value',
textfont_size=20,
                  marker=dict(line=dict(color='#000000', width=3)))
fig.update_layout(title="Most popular cuisines of Bangalore
",
                  titlefont={'size': 30},
                  )
fig.show()
```



Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

```
MughlaiFoodcafe = data[data['cuisines'].str.contains('Mughlai',
case=False, regex=True,na=False)]
MughlaiFoodcafe.head()

#pie chart showing % of various Food serving Type cafe
slices=[MughlaiFoodcafe.shape[0],
        ChineseFoodcafe.shape[0],
        MexicanFoodcafe.shape[0],
        NorthIndianFoodcafe.shape[0],
        SouthIndianFoodcafe.shape[0],
        ItalianFoodcafe.shape[0],
        AmericanFoodcafe.shape[0]]
labels=['Mughlai','Chinese','Mexican','North Indian','South
Indian','Italian','American']
colors = ['#3333cc','#ffff1a','#ff3333','#c2c2d6','#6699ff','#c4ff4d','#339933']
plt.pie(slices,colors=colors, labels=labels, autopct='%1.0f%%',
pctdistance=.5, labeldistance=1.2,shadow=True)
fig = plt.gcf()
plt.title("Percentage of cafe according to their Food Type",
bbox={'facecolor':'1', 'pad':5})
fig.set_size_inches(12,12)
plt.show()
```



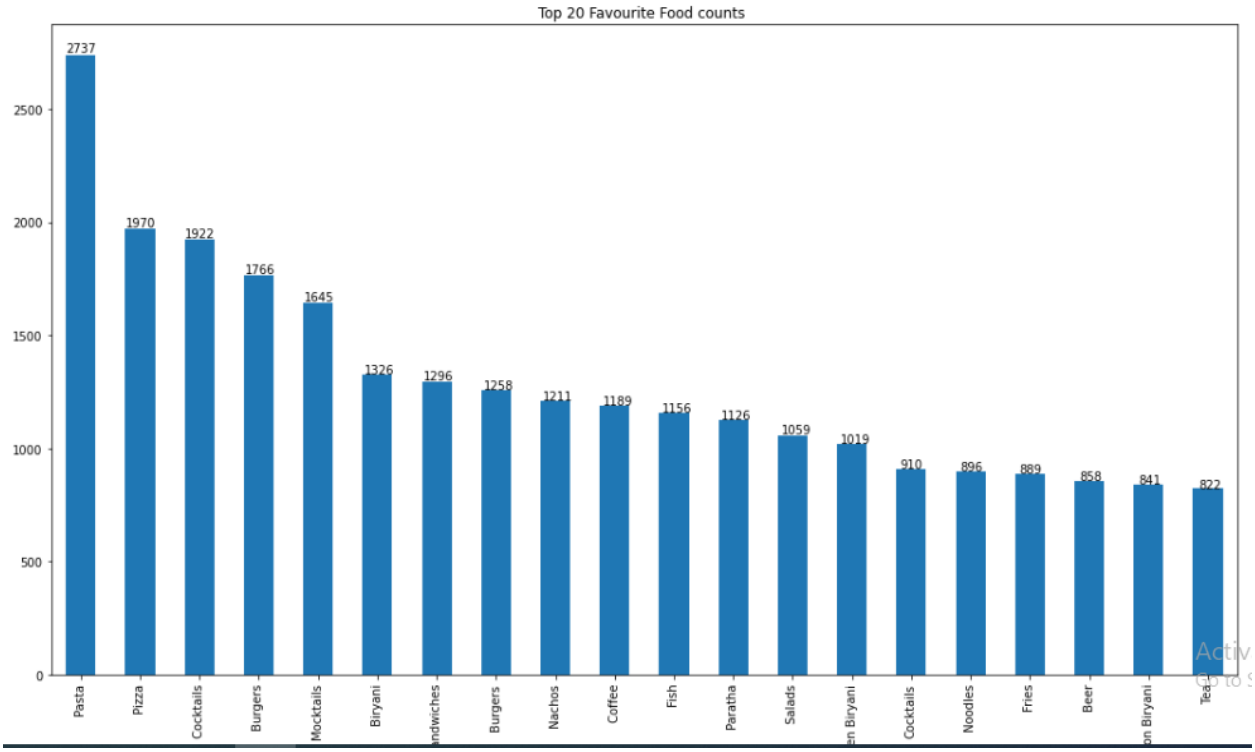
Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

```
# Most Liked Dishes in Bangalore
import re
data=data[data['dish_liked'].notnull()]
data.index=range(data.shape[0])
likes=[]
for i in range(data.shape[0]):
    splited_array=re.split(',',data['dish_liked'][i])
    for item in splited_array:
        likes.append(item)

print("Count of Most liked dishes of Bangalore")
favourite_food = pd.Series(likes).value_counts()
favourite_food.head(20)
```

```
Count of Most liked dishes of Bangalore
 Pasta                2737
 Pizza                1970
 Cocktails            1922
 Burgers              1766
 Mocktails            1645
 Biryani              1326
 Sandwiches           1296
Burgers               1258
 Nachos               1211
 Coffee               1189
 Fish                 1156
 Paratha              1126
 Salads               1059
 Chicken Biryani      1019
Cocktails              910
 Noodles               896
 Fries                 889
 Beer                  858
 Mutton Biryani        841
 Tea                   822
```

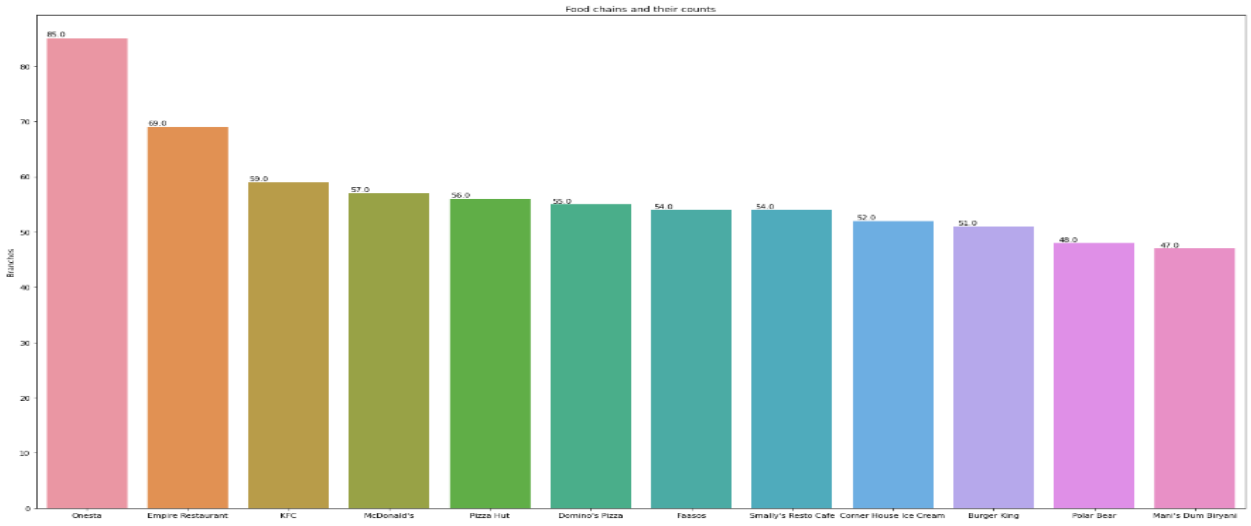
```
ax = favourite_food.nlargest(n=20,
keep='first').plot(kind='bar',figsize=(15,15),title = 'Top 20
Favourite Food counts ')
for p in ax.patches:

   ax.annotate(str(p.get_height()), (p.get_x() * 1.005,
p.get_height() * 1.005))
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu




```
#Analysis of biggest food chains
branches =
data.groupby(['name']).size().to_frame('count').reset_index().
sort_values(['count'],ascending=False)
ax = sns.barplot(x='name', y='count', data=branches[:12])
plt.xlabel('')
plt.ylabel('Branches')
plt.title('Food chains and their counts')
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.005,
p.get_height() * 1.005))

fig = plt.gcf()
fig.set_size_inches(25,15)
```

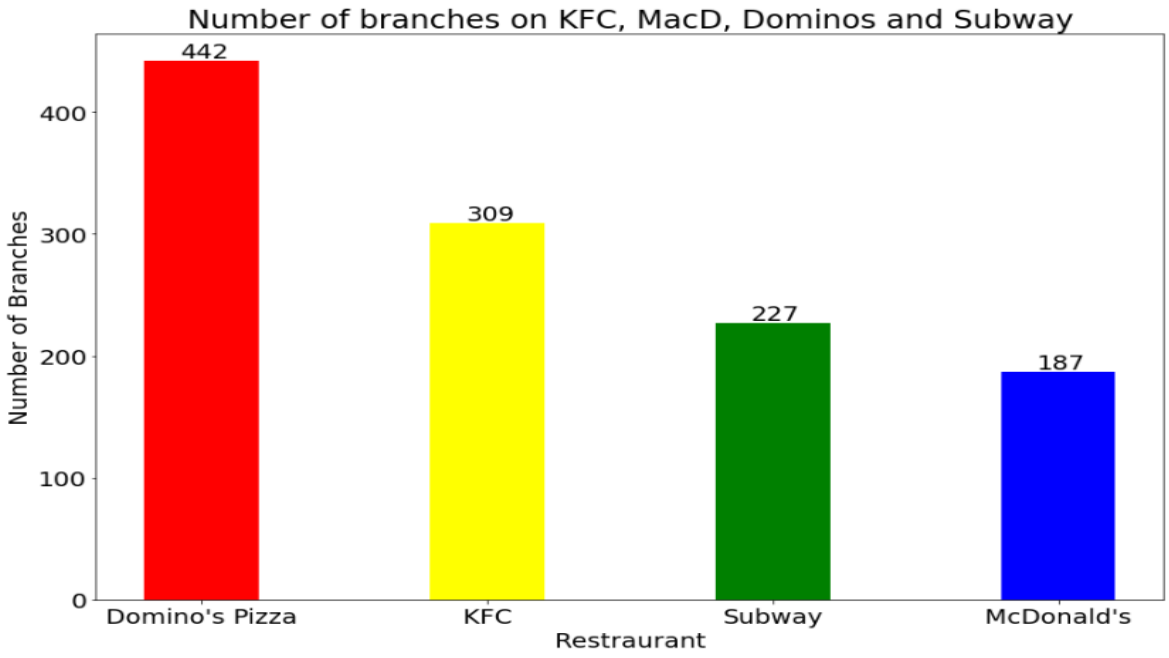
Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu



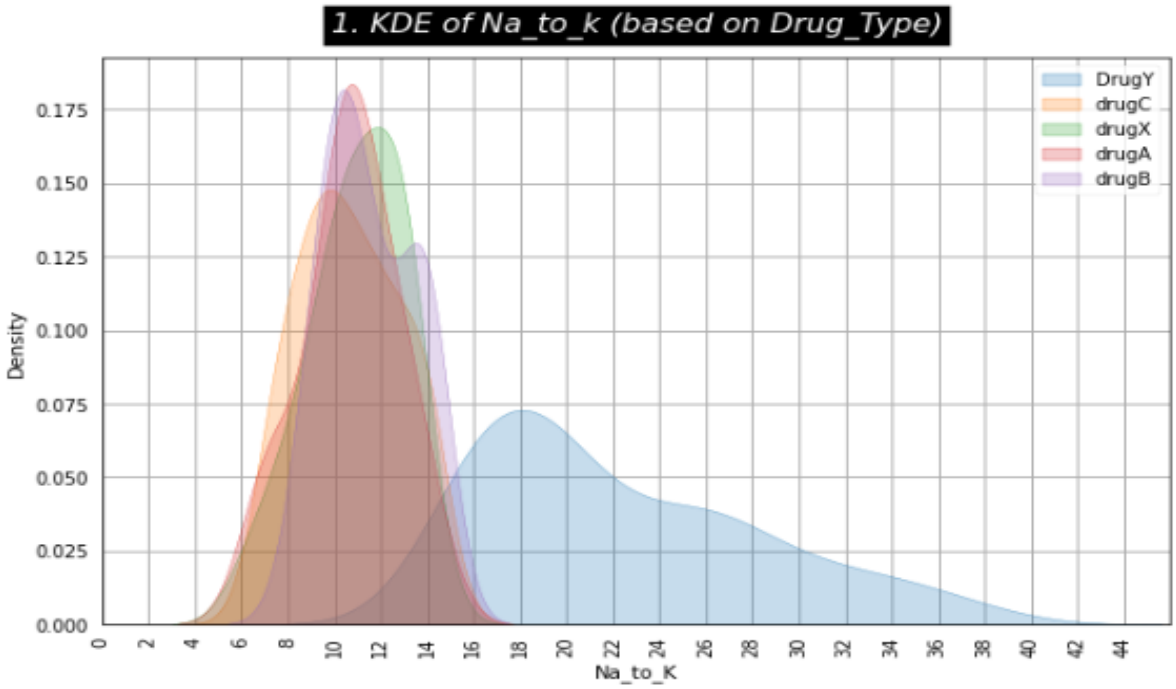

```
data = df.groupby('name').size()[["Domino's Pizza", "KFC",
"McDonald's",'Subway']].sort_values(ascending = False)

x    = data.index
y    = data.values

plt.figure(figsize = (15,10))

color = ['red','yellow','green','blue']
ax = plt.bar(x,y,width = 0.4,color = color)
for i in ax:
    x_ = i.xy[0] + i.get_width() / 2
    y_ = i.get_height()
    txt = str(y_)
    plt.annotate(
        text       = txt,
        xy         = (x_,y_),
        xytext     = (-17,2.9),
        textcoords = 'offset points'
    )
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

```
plt.xlabel('Restraurant')
plt.ylabel('Number of Branches')
plt.title('Number of branches on KFC, MacD, Dominos and Subway')
plt.show()
```

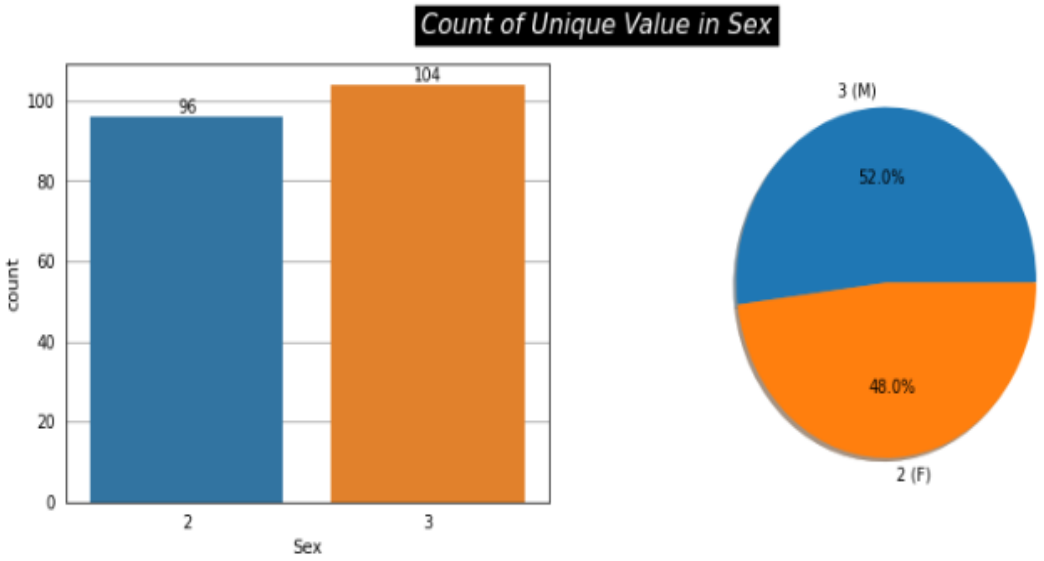


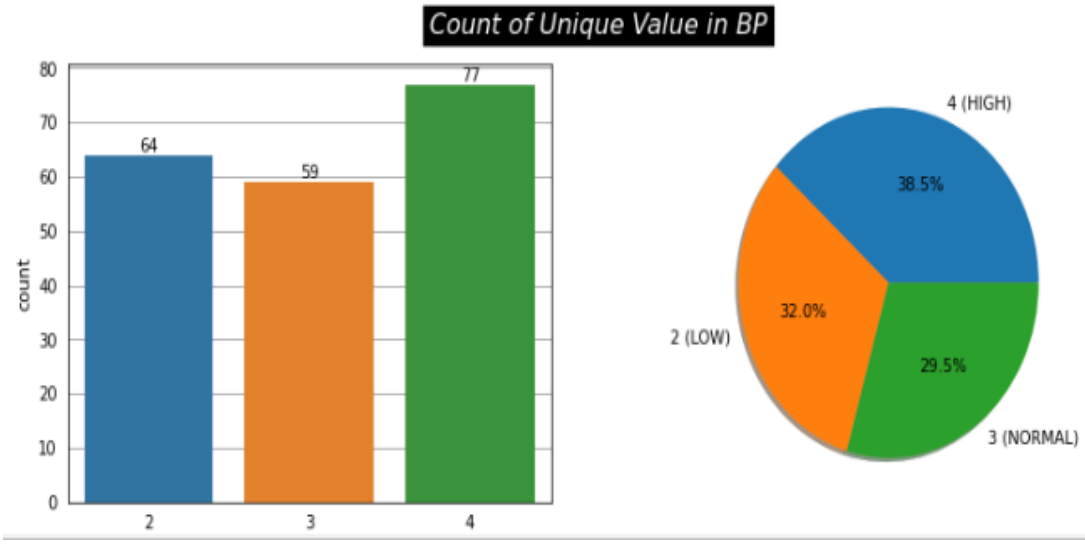

```
plt.style.use('seaborn-notebook')
for i, label in enumerate(df.Drug_Type.unique().tolist()):
    sns.kdeplot(df2.loc[df2['Drug_Type'] == i+1, 'Na_to_K'], label=label, shade=True)

plt.title('1. KDE of Na_to_k (based on Drug_Type)', fontdict=font, pad=15)
plt.xticks(np.arange(0,46,2), rotation=90)
plt.xlim([0,46])
plt.legend()
plt.show()
```




```
# draw countplot and pie plot of categorical data
for col in categorical:
    fig, axes = plt.subplots(1,2,figsize=(10,4))
    # count of col (countplot)
    sns.countplot(data=df2, x=col, ax=axes[0])
    for container in axes[0].containers:
        axes[0].bar_label(container)
    # count of col (pie chart)
    slices = df2[col].value_counts().values
    activities = [f"{i} ({var})" for i, var in
zip(df2[col].value_counts().index,
df[col].value_counts().index)]
    axes[1].pie(slices, labels=activities, shadow=True,
autopct='%1.1f%%')
    plt.suptitle(f'Count of Unique Value in {col}', y=1.09,
**font)
    plt.show()
```

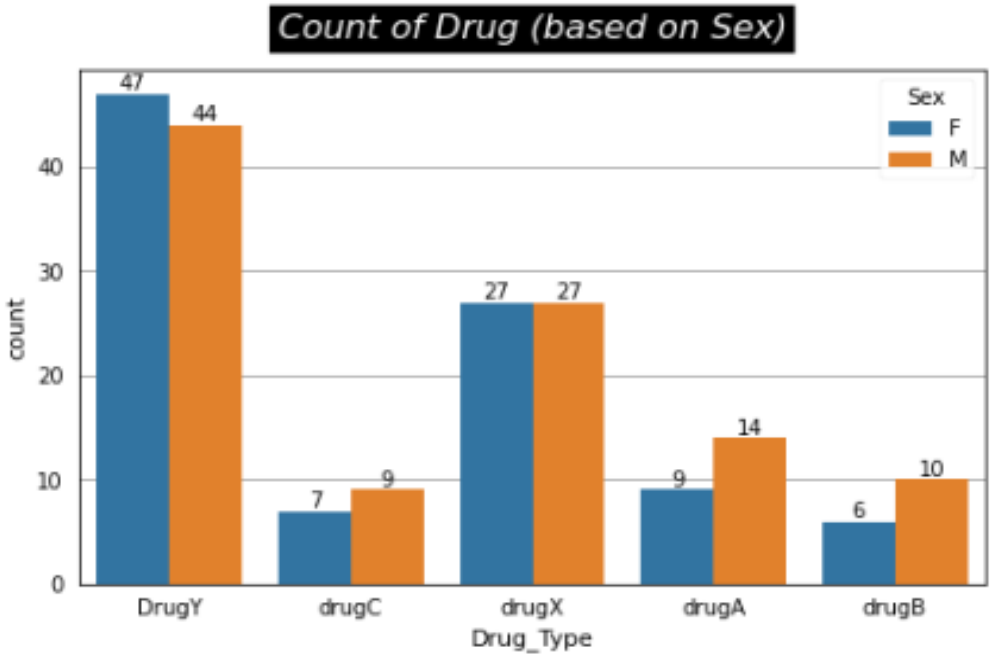
Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu






```
for col in ['Sex','BP','Cholesterol']:
    ax = sns.countplot(data=df, x='Drug_Type', hue=col)
    for container in ax.containers:
        ax.bar_label(container)
    plt.title(f'Count of Drug (based on {col})', fontdict=font,
pad=15)
    plt.show()
```

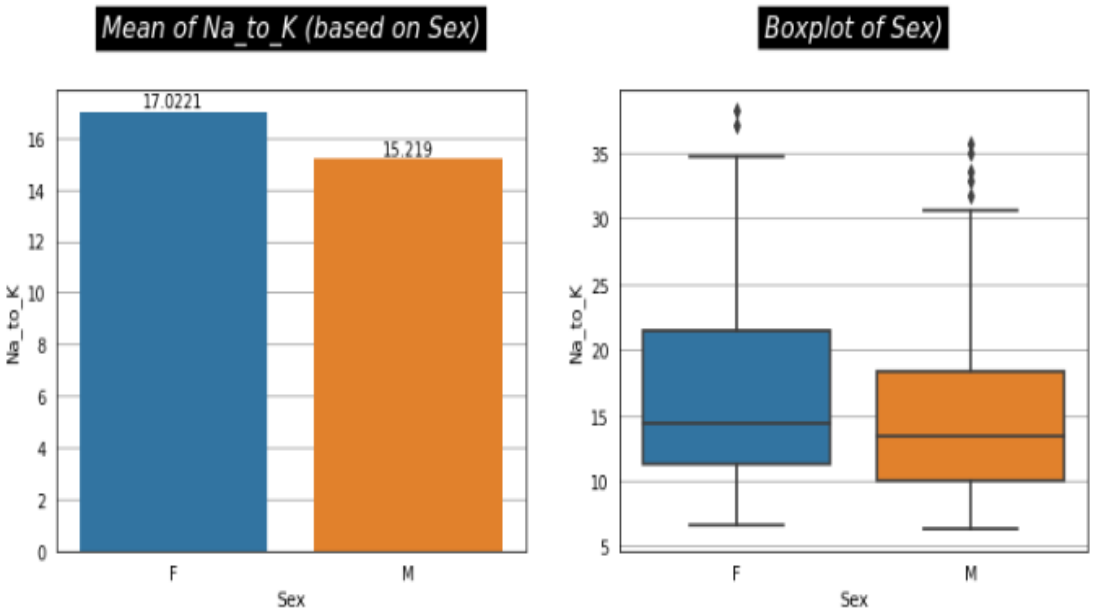
Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu



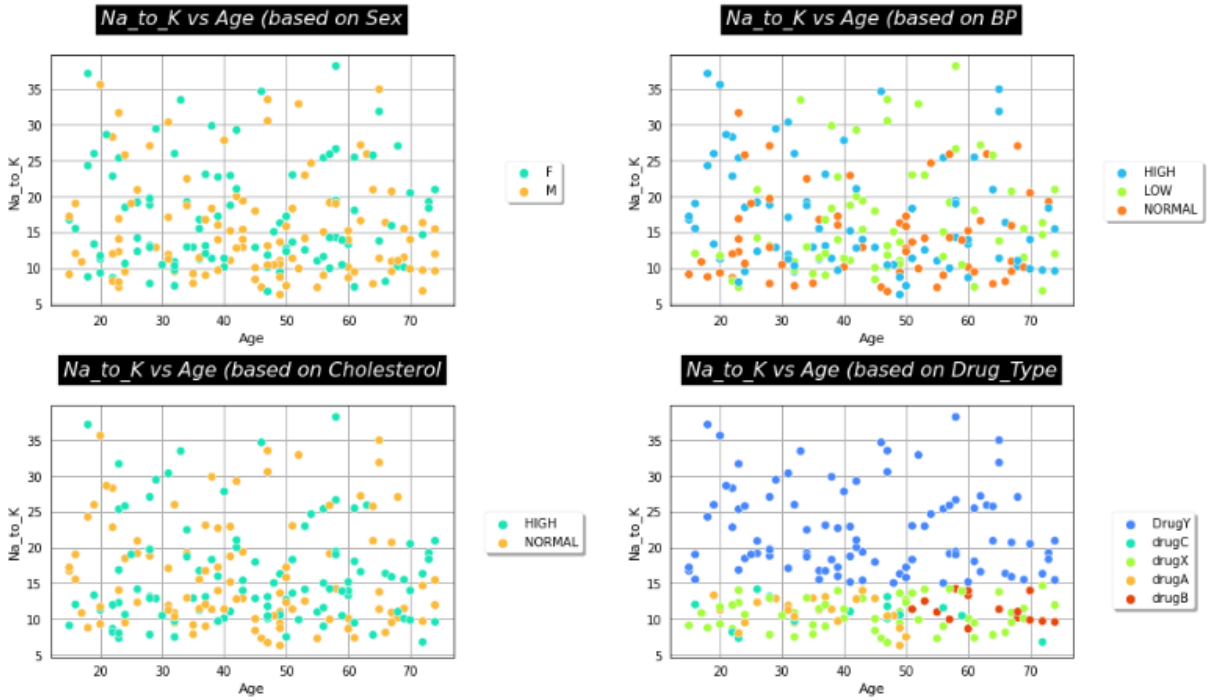

```
for col in ['Sex', 'BP', 'Cholesterol']:
    fig , ax= plt.subplots(1,2, figsize=(10,4))
    gp =
df.groupby([col])['Na_to_K'].mean().to_frame().reset_index()
    sns.barplot(data=gp, x=col, y='Na_to_K', ax=ax[0])
    for container in ax[0].containers:
        ax[0].bar_label(container)
    ax[0].set_title(f'Mean of Na_to_K (based on {col})', y=1.09,
**font)
    sns.boxplot(data=df, x=col, y='Na_to_K', ax=ax[1])
    ax[1].set_title(f'Boxplot of {col})', y=1.09, **font)
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

```
    plt.show()
```

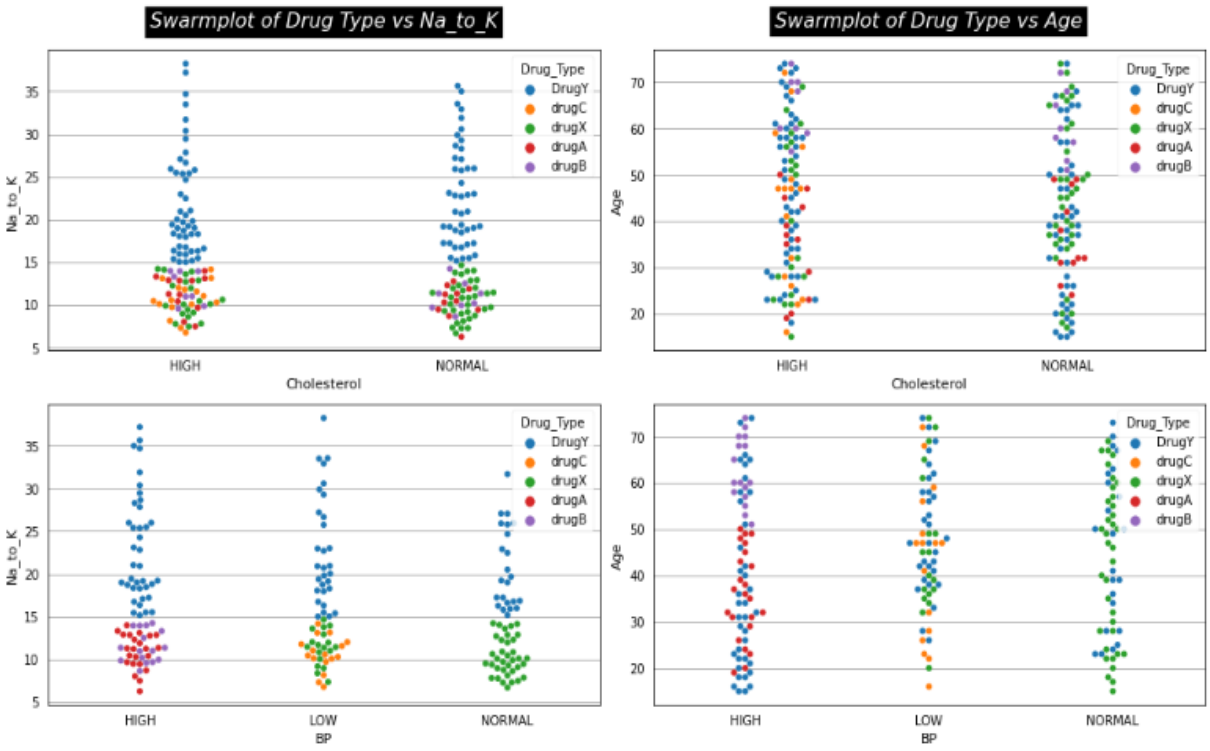



```
fig, ax = plt.subplots(2,2,figsize=(14,8))
for i, col in enumerate(['Sex', 'BP', 'Cholesterol',
'Drug_Type']):
    sns.scatterplot(data=df, x='Age', y='Na_to_K', hue=col,
ax=ax[i//2, i%2], palette='turbo')
    ax[i//2, i%2].set_title(f'Na_to_K vs Age (based on {col}',
y=1.09, **font)
    ax[i//2, i%2].legend(loc='upper center',
bbox_to_anchor=(1.2, 0.6),
        fancybox=True, shadow=True)

fig.tight_layout()
plt.show()
```




```
fig, ax = plt.subplots(3,2,figsize=(14,12))
sns.swarmplot(data=df, x='Cholesterol', y='Na_to_K',
hue='Drug_Type', ax=ax[0,0])
sns.swarmplot(data=df, x='Cholesterol', y='Age',
hue='Drug_Type', ax=ax[0,1])
sns.swarmplot(data=df, x='BP', y='Na_to_K',
hue='Drug_Type', ax=ax[1,0])
sns.swarmplot(data=df, x='BP', y='Age', hue='Drug_Type',
ax=ax[1,1])
sns.swarmplot(data=df, x='Sex', y='Na_to_K',
hue='Drug_Type', ax=ax[2,0])
sns.swarmplot(data=df, x='Sex', y='Age', hue='Drug_Type',
ax=ax[2,1])
ax[0,0].set_title('Swarmplot of Drug Type vs Na_to_K',y=1.05,
**font)
ax[0,1].set_title('Swarmplot of Drug Type vs Age',y=1.05,
**font)
plt.tight_layout()
```

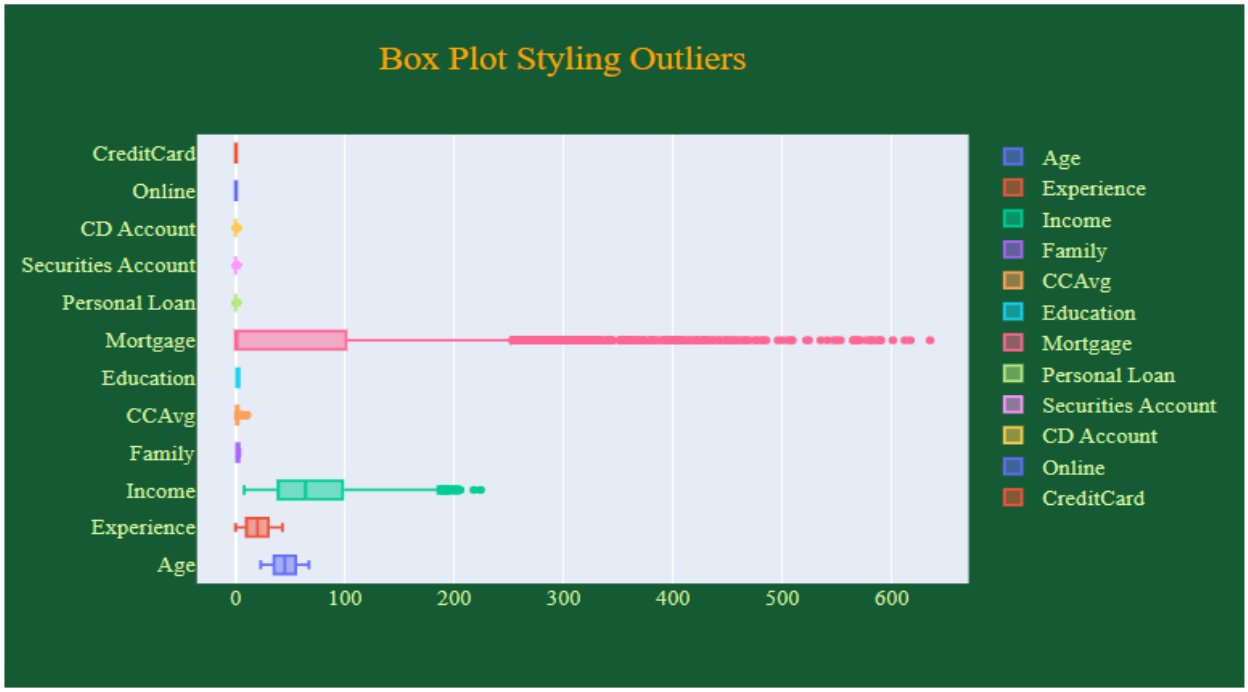
Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu




```
import itertools
import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
from plotly.subplots import make_subplots
# for solve problem of show plotly plots
from plotly.offline import init_notebook_mode
init_notebook_mode(connected=True)
plt.style.use('_mpl-gallery')
FONT = {'fontsize':20, 'fontstyle':'normal', 'fontfamily':'Times
New Roman', 'backgroundcolor':'#145A32', 'color':'orange'} #
for plot title

fig = go.Figure()
for col in df:
    fig.add_trace(go.Box(x=df[col], name=col))
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

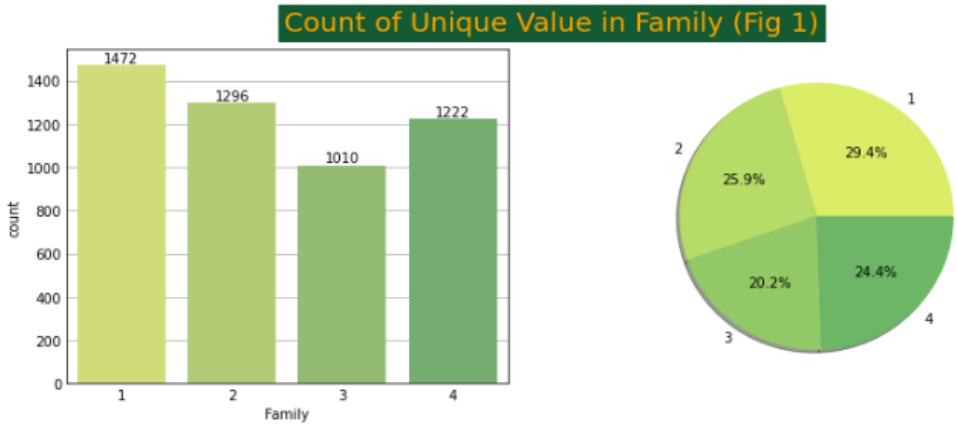
```
fig.update_layout(
    title_text="Box Plot Styling Outliers",
    title_font=dict(color='orange', family='newtimeroman',
size=25),
    title_x=0.45,
    paper_bgcolor='#145A32',
    # plot_bgcolor='#DAF7A6',
    font=dict(color='#DAF7A6', family='newtimeroman',
size=16),
    )
fig.show()
```

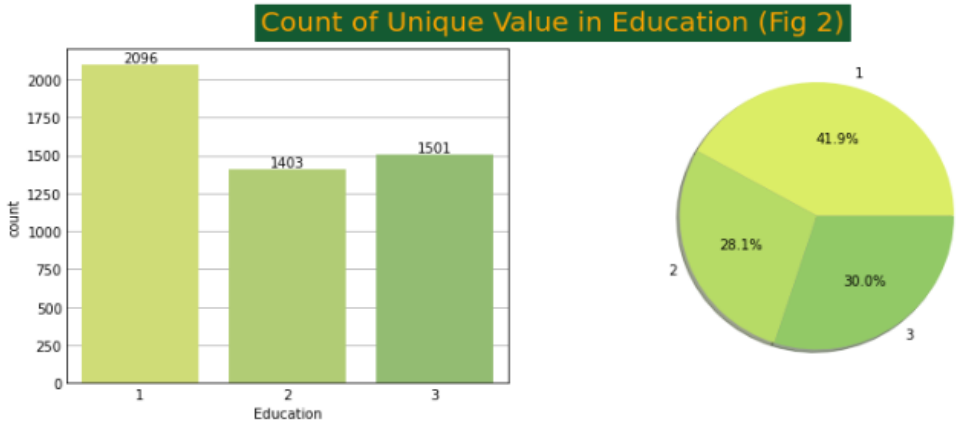



```
# univariate analysis of categorical data:
sns.set_palette("summer_r")
for i, col in enumerate(discrete_cols1):

    fig, axes = plt.subplots(1,2,figsize=(10,4))
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

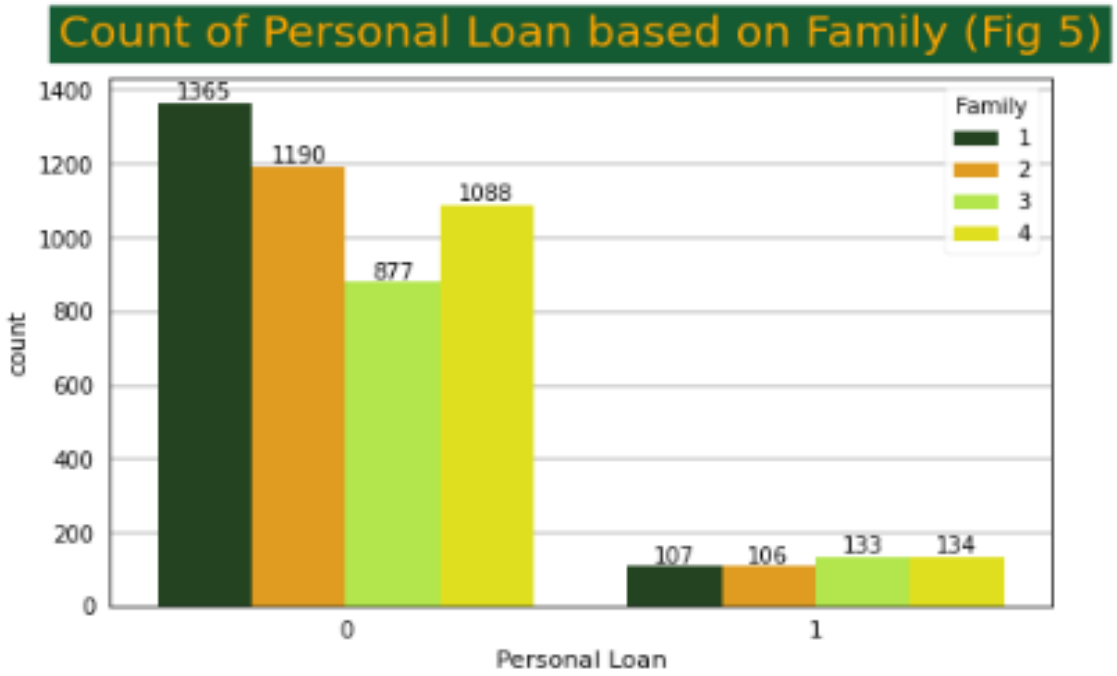
```
    # count of col (countplot)
    sns.countplot(data=df, x=col, ax=axes[0])
    for container in axes[0].containers:
        axes[0].bar_label(container)
    # count of col (pie chart)
    slices = df[col].value_counts().sort_index().values
    activities = [var for var in
df[col].value_counts().sort_index().index]
    axes[1].pie(slices, labels=activities, shadow=True,
autopct='%1.1f%%')
    plt.suptitle(f'Count of Unique Value in {col} (Fig {i+1})',
y=1.09, **FONT)
    plt.show()
```





Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

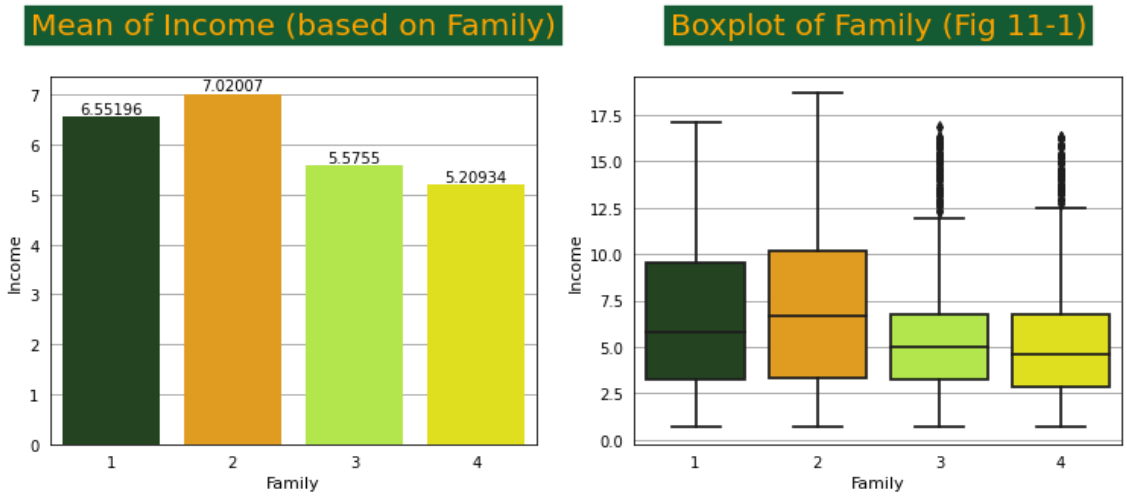
```
sns.set_palette(['#1f4a1b','orange','#bbff33','yellow'])
discrete_cols2 = ['Family', 'Education', 'Securities Account',
'CD Account', 'Online', 'CreditCard']
for i, col in enumerate(discrete_cols2):
    ax = sns.countplot(data=df, x='Personal Loan', hue=col)
    for container in ax.containers:
        ax.bar_label(container)
    plt.title(f'Count of Personal Loan based on {col} (Fig {i+5})',
fontdict=FONT, pad=15)
    plt.show()
```




```
for i, col in enumerate(['Income', 'CCAvg','Mortgage']):
    print('='*30, f"Mean of {col} in each categorical feature",
'='*30)
    for j, cat in enumerate(discrete_cols2):
        fig , ax= plt.subplots(1,2, figsize=(10,4))
        gp =
df.groupby([cat])[col].mean().to_frame().reset_index()
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

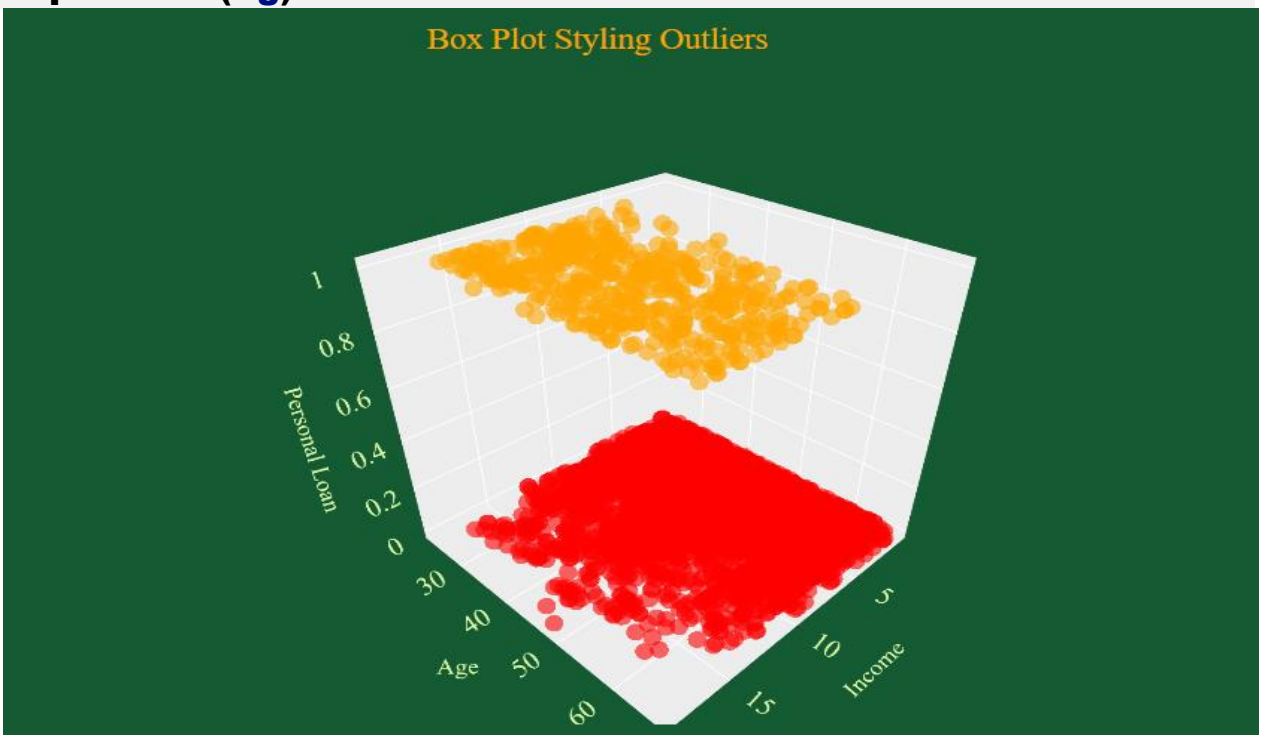
```
        sns.barplot(data=gp, x=cat, y=col, ax=ax[0])
        for container in ax[0].containers:
            ax[0].bar_label(container)
        ax[0].set_title(f'Mean of {col} (based on {cat})', y=1.09,
**FONT)

        sns.boxplot(data=df, x=cat, y=col, ax=ax[1])
        ax[1].set_title(f'Boxplot of {cat} (Fig {i+11}-{j+1})',
y=1.09, **FONT)

        plt.show()
```




```
continuous_cols = ['Age','Experience','CCAvg','Mortgage']

for i, col in enumerate(continuous_cols):
    fig = px.scatter_3d(
        data_frame= df,
        x=df.Income,
        y=df[col],
        z=df['Personal Loan'],
        color=df['Personal Loan'].astype(str),
        color_discrete_map={'1':'orange', '0':'red'},
        template='ggplot2',
```

Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu

```
        hover_name='Age',
        # hover_data=
        opacity=0.6,
        # symbol='Transmission',
        # symbol_map=
        # log_x=True,
        # log_z=True,
        height=700,
        title=f'3D scatter of features based on Personal Loan (Fig
{i+1})')
    fig.update_layout(
        title_text="Box Plot Styling Outliers",
        title_font=dict(color='orange', family='newtimeroman',
size=25),
        title_x=0.45,
        paper_bgcolor='#145A32',
        # plot_bgcolor='#DAF7A6',
        font=dict(color='#DAF7A6', family='newtimeroman', size=16),
    )
    pio.show(fig)
```



Syed Afroz Ali (Kaggle Grandmaster)
Follow me on LinkedIn for more: https://lnkd.in/g3W7w_Xu