In [1]:
```
import pandas as pd
import numpy as np

df= pd.read_csv ('covidReactions_Jan_Aug.csv')

C:\Users\Deepyamaneedhi\lib\site-packages\IPython\core\interactiveshell.py:3165: DtypeWarning: Columns (24,29,32,40) have mixed types.Specify dtype option on import o
r set low_memory=False.
  raw_cells = self.shell.transform_cell(cell.source,kw_code_cell,kw_cell_name,

# Deleting the column needed
df_covid = df[['VAERS_ID', 'VAX_MANU', 'AGE_YRS', 'SEX', 'SYMPTOM1', 'SYMPTOM2', 'SYMPTOM3', 'SYMPTOM4', 'SYMPTOM5', 'HOSPDAYS', 'HOSPITAL', 'VAX_DOSE_SERIES']]
df_covid.head()
```

Out[1]:

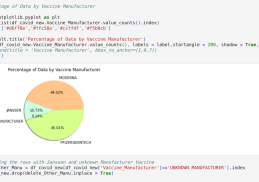
	VAERS_ID	VAX_MANU	AGE_YRS	SEX	SYMPTOM1	SYMPTOM2	SYMPTOM3	SYMPTOM4	SYMPTOM5	HOSPDAYS	HOSPITAL	VAX_DOSE_SERIES
0	916600	MODERNA	33.0	F	Dysphagia	Epiglottitis	NaN	NaN	NaN	NaN	NaN	1
1	916601	MODERNA	73.0	F	Anxiety	Dyspnoea	NaN	NaN	NaN	NaN	NaN	1
2	916602	PFIZER\BIONTECH	23.0	F	Chest discomfort	Dysphagia	Pain in extremity	Visual impairment	NaN	NaN	NaN	1
3	916603	MODERNA	58.0	F	Dizziness	Fatigue	Mobility decreased	NaN	NaN	NaN	NaN	UNK
4	916604	MODERNA	47.0	F	Injection site erythema	Injection site pruritus	Injection site swelling	Injection site warmth	NaN	NaN	NaN	1
...	...	...	...	...	...	...	...	...	...	...	...	...
603457	1514865	PFIZER\BIONTECH	33.0	F	Lethargy	Malaise	Nausea	Seizure	Syncope	2.0	Y	1
603458	1514863	PFIZER\BIONTECH	31.0	F	Unresponsive to stimuli	Vomiting	NaN	NaN	NaN	2.0	Y	1
603459	1514847	PFIZER\BIONTECH	28.0	F	Incorrect dose administered	NaN	NaN	NaN	NaN	NaN	NaN	1
603460	1514868	PFIZER\BIONTECH	23.0	F	Incorrect product formulation administered	NaN	NaN	NaN	NaN	NaN	NaN	1
603461	1514855	PFIZER\BIONTECH	25.0	F	Hypotension	Syncope	Unresponsive to stimuli	NaN	NaN	NaN	NaN	1

600294 rows × 12 columns

In [4]:
```
#changed the names
df_covid.rename(columns={'VAX_MANU': 'Vaccine Manufacturer', 'AGE_YRS': 'Age', 'SEX': 'Gender', 'SYMPTOM1': 'Symptom1', 'SYMPTOM2': 'Symptom2', 'SYMPTOM3': 'Symp

# the count of vaccine manufacturers
df_covid.nunique()['Vaccine Manufacturer'].value_counts()
```

Out[5]:
```
PFIZER\BIONTECH    271633
MODERNA            265652
JANSSEN             58192
UNKNOWN MANUFACTURER     1824
Name: Vaccine Manufacturer, dtype: int64
```

In [6]:
```
# Percentage of Data by Vaccine Manufacturer

import matplotlib.pyplot as plt
label = list(df_covid.nunique()['Vaccine Manufacturer'].value_counts().index)
colour = ['#ffc7ce', '#ffcf8a', '#cfcf6f', '#f5b4b3']

title = plt.title('Percentage of Data by Vaccine Manufacturer')
plt.pie(df_covid.nunique()['Vaccine Manufacturer'].value_counts(), labels = label,startangle = 200, shadow = True,colors = colour,autopct='%1.2f%%')
plt.legend(title = 'Vaccine Manufacturer', bbox_to_anchor=(1,0.7))
plt.show()
```



In [7]:
```
# deleting the rows with Janssen and unknown Manufacturer Vaccine
delete_Other_Manu = df_covid.index[df_covid.nunique()['Vaccine Manufacturer']=='UNKNOWN MANUFACTURER'].index
df_covid.drop(delete_Other_Manu,inplace = True)

delete_Janssen = df_covid.index[df_covid.nunique()['Vaccine Manufacturer']=='JANSSEN'].index
df_covid.drop(delete_Janssen,inplace = True)

#checked the count of vaccines
df_covid.nunique()['Vaccine Manufacturer'].value_counts()
```

Out[7]:
```
PFIZER\BIONTECH    271633
MODERNA            265652
Name: Vaccine Manufacturer, dtype: int64
```

In [8]:
```
#drop rows with null values in age
df_covid.nunique()['Age'].isnull = True
df_covid.dropna(subset=['Age'],inplace = True)

#drop rows containing dose 2 only
delete_dose2 = df_covid.index[df_covid.nunique()['MAX_DOSE_SERIES']=='2'].index
df_covid.drop(delete_dose2,inplace = True)

# separate according to age groups
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>20) & (df_covid.nunique()['Age']<=30)] = '10-30'
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>=30) & (df_covid.nunique()['Age']<=40)] = '30-40'
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>=40) & (df_covid.nunique()['Age']<=50)] = '40-50'
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>=50) & (df_covid.nunique()['Age']<=60)] = '50-60'
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>=60) & (df_covid.nunique()['Age']<=70)] = '60-70'
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>=70) & (df_covid.nunique()['Age']<=80)] = '70-80'
df_covid.nunique()['Age'].loc[(df_covid.nunique()['Age']>=80)] = '80+'

print(df_covid.nunique()['Age_Group'].value_counts())
```

```
50-60    98727
40-50    87304
30-40    87004
60-70    66358
10-30    57767
70-80    34139
80+      12335
Name: Age_Group, dtype: int64
```

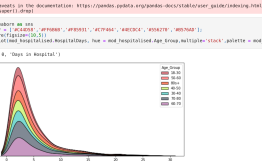
In [9]:
```
df_Moderna = df_covid.nunique()['Vaccine Manufacturer'] == 'MODERNA'].copy()
df_Pfizer = df_covid.nunique()['Vaccine Manufacturer'] == 'PFIZER\BIONTECH'].copy()

#delete rows with unknown age
null_age = df_Moderna[df_Moderna.Age == 'Unknown'].index
df_Moderna.drop(null_age,inplace= True)

pfizer_null_age = df_Pfizer[df_Pfizer.Age == 'Unknown'].index
df_Pfizer.drop(pfizer_null_age,inplace= True)
```

In [10]:
```
moderna_Symptom = df_Moderna.Symptom1.value_counts().sort_values(ascending=False).head(10)
moderna_Symptom = moderna_Symptom.sort_values()
moderna_color = ['#ffd8b0', '#d3f5ce', '#fcf7bc', '#f3b4c1', '#bdf0ae', '#f3b6c9', '#cfcf6f', '#ffc5da', '#d8ff8e', '#ffd6b1']

plt.figure(figsize=(5,4))
plt.barh(moderna_Symptom.index,moderna_Symptom.values,color=moderna_color )
plt.title('Most common Adverse reactions caused by the Moderna and Pfizer Vaccine')
plt.ylabel('Moderna Vaccine')
for index, value in enumerate(moderna_Symptom.values):
    plt.text(value, index, str(value))
plt.show()
```

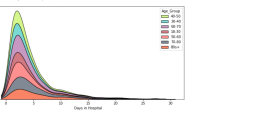
In [11]:
```
pfizer_Symptom = df_Pfizer.Symptom1.value_counts().sort_values(ascending=False).head(10)
pfizer_Symptom = pfizer_Symptom.sort_values()
pfizer_color = ['#ffd8b0', '#d3f5ce', '#fcf7bc', '#f3b4c1', '#bdf0ae', '#f3b6c9', '#cfcf6f', '#ffc5da', '#d8ff8e', '#ffd6b1']

plt.figure(figsize=(5,4))
plt.barh(pfizer_Symptom.index,pfizer_Symptom.values,color=pfizer_color)
plt.title('Most common Adverse reactions caused by the Moderna and Pfizer Vaccine')
plt.ylabel('Pfizer Vaccine')

for index, value in enumerate(pfizer_Symptom.values):
    plt.text(value, index, str(value))

plt.show()
```

In [13]:
```
import matplotlib.pyplot as plt
import numpy as np
#chance of getting a reaction from the pfizer
y = np.array([273633,1875341])
mylabels = ["Reactions(<0.70%)", "No Reactions(>99.25%)"]
plt.figure(figsize=(5,5))
myexplode = [0.4,0]

plt.pie(y,labels = mylabels,explode = myexplode,startangle = 150,colors = ['#8D000D', '#e5f6d5'])
plt.show()
```

In [14]:
```
#chance of getting a reaction from the Moderna
y = np.array([265052,76318253])
mylabels = ["Reactions(<0.35%)", "No Reactions(>99.65%)"]
plt.figure(figsize=(5,5))
myexplode = [0.4,0]

plt.pie(y,labels = mylabels,explode = myexplode,startangle = 150,colors = ['#8D000D', '#e5f6d5'])
plt.show()
```



In [15]:
```
#no. of hospitalisations after an adverse reaction to MODERNA
\hls = 'Hospitalised', 'Not Hospitalised'
mod_hospitalised = round(list(df_Moderna.Hospital.value_counts()).values/df_Moderna.shape[0]*100)[0],2)
plt.figure(figsize=(5,5))
valx= [mod_hospitalised,mod_hospitalised,]
myexplode = [0.0,2]
plt.pie(valx,labels= hls,startangle = 360,explode = myexplode,autopct = '%1.2f%%', colors = ['#d6ff8e', '#e5f6d5'])
plt.show()
```



In [17]:
```
#no. of hospitalised after an adverse reaction - PFIZER
pfizer_hospitalised = round(list(df_Pfizer.Hospital.value_counts()).values/df_Pfizer.shape[0])*100)[0],2)
pfizer_hospital_no = 100 - pfizer_hospitalised
valx= [pfizer_hospitalised,pfizer_hospital_no]
myexplode = [0.0,2]
plt.pie(pfizer_valx,labels= hls,startangle = 360,explode = myexplode,autopct = '%1.2f%%', colors = ['#8dffc8', '#ed9ce4'])
plt.show
```

Out[17]:
```
<function matplotlib.pyplot.show(*args,*kwargs)>
```


In [20]:
```
#create new dataframe
mod_hospitalised = df_Moderna[~df_Moderna.Hospital.isna()]
pfizer_hospitalised = df_Pfizer[~df_Pfizer.Hospital.isna()]

#delete null values in date
#delete rows with more than 30 days
mod_null_hospital = mod_hospitalised[mod_hospitalised.HospitalDays>=30].index
mod_hospitalised.drop(mod_null_hospital,inplace = True)

pfizer_null_hospital = pfizer_hospitalised[pfizer_hospitalised.HospitalDays>=30].index
pfizer_hospitalised.drop(pfizer_null_hospital,inplace = True)
```

```
C:\Users\Deepyamaneedhi\lib\site-packages\pandas\core\frame.py:4308: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(
C:\Users\Deepyamaneedhi\lib\site-packages\pandas\core\frame.py:4308: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(
```

In [21]:
```
import seaborn as sns
mod_color = ['#c4d6b0', '#FF8080', '#F0E051', '#C1FF46', '#40CEC4', '#95A579', '#B57BA8']
sns.set_palette(mod_color)
sns.kdeplot(mod_hospitalised.HospitalDays, hue = mod_hospitalised.Age_Group,multiple='stack',palette = mod_color).set_xlabel('Days in Hospital')
```

Out[21]:
```
Text(0.5, 0, 'Days in Hospital')
```


In [22]:
```
pfizer_color = ['#C7F464', '#4ECDC4', '#B57BA8', '#C4A050', '#FF6DBD', '#55C570', '#FFD531'];
plt.figure(figsize=(30,5))
sns.kdeplot(pfizer_hospitalised.HospitalDays, hue = pfizer_hospitalised.Age_Group,multiple='stack',palette = pfizer_color).set_xlabel('Days in Hospital')
```

Out[22]:
```
Text(0.5, 0, 'Days in Hospital')
```