```
In [1]:  import pandas as pd
         import seaborn as sns
         import matplotlib.pyplot as plt
         from datetime import datetime
         import plotly.express as px
         import plotly.graph_objects as go

         'Access the code for this script at https://github.com/alejandromignucci/AI395T'
```

Out[1]: 'Access the code for this script at https://github.com/alejandromignucci/AI395T'

```
In [2]:  patients = pd.read_csv('patients.csv')

         def calculate_age(dob, dischtime):
             dob = datetime.strptime(str(dob), "%Y-%m-%d %H:%M:%S")
             disc = datetime.strptime(str(dischtime), '%Y-%m-%d %H:%M:%S')
             return (disc - dob).days // 365


         admissions = pd.read_csv('admissions.csv')

         merged_data = pd.merge(admissions, patients, on='SUBJECT_ID')

         merged_data['age'] = merged_data.apply(lambda x: calculate_age(x['DOB'], x['DISCHTIME']), axis=1)
         merged_data = merged_data[merged_data['age'] < 120]

         merged_data.head()
```

Out[2]:

	ROW_ID_x	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURANCE
0	21	22	165315	2196-04-09 12:26:00	2196-04-10 15:54:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	DISC-TRAN CANCER/CHLDRN H	Private
1	22	23	152223	2153-09-03 07:15:00	2153-09-08 19:10:00	NaN	ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Medicare
2	23	23	124321	2157-10-18 19:34:00	2157-10-25 14:00:00	NaN	EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME HEALTH CARE	Medicare
3	24	24	161859	2139-06-06 16:14:00	2139-06-09 12:48:00	NaN	EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME	Private
4	25	25	129635	2160-11-02 02:06:00	2160-11-05 14:55:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME	Private

5 rows × 27 columns

```
In [3]:  merged_data.count()
```

Out[3]: ROW_ID_x   56360
         SUBJECT_ID   56360
         HADM_ID   56360
         ADMITTIME   56360
         DISCHTIME   56360
         DEATHTIME   5305
         ADMISSION_TYPE   56360
         ADMISSION_LOCATION   56360
         DISCHARGE_LOCATION   56360
         INSURANCE   56360
         LANGUAGE   31892
         RELIGION   55921
         MARITAL_STATUS   46373
         ETHNICITY   56360
         EDREGTIME   28692
         EDOUTTIME   28692
         DIAGNOSIS   56338
         HOSPITAL_EXPIRE_FLAG 56360
         HAS_CHARTEVENTS_DATA 56360
         ROW_ID_y   56360
         GENDER   56360
         DOB   56360
         DOD   20598
         DOD_HOSP   13794
         DOD_SSN   17402
         EXPIRE_FLAG   56360
         age   56360
         dtype: int64

```python
In [4]:  #Print the dataframe for one single patient to check to see the amount of admissions over the years
         merged_data[merged_data['SUBJECT_ID'] == 188].head(10)
```

Out[4]:

	ROW_ID_x	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURAI
465	245	188	160697	2157-01-11 16:56:00	2157-01-19 14:58:00	NaN	EMERGENCY	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Pri
466	246	188	191517	2157-03-07 11:08:00	2157-03-10 13:50:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME	Pri
467	247	188	150463	2157-11-17 12:11:00	2157-11-20 13:05:00	NaN	EMERGENCY	CLINIC REFERRAL/PREMATURE	HOME	Pri
468	248	188	192557	2160-11-25 21:55:00	2160-11-28 12:42:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Pri
469	249	188	123860	2160-12-31 12:34:00	2161-01-08 19:50:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Pri
470	250	188	164735	2161-07-01 19:44:00	2161-07-10 20:00:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Pri
471	251	188	132401	2161-11-01 17:48:00	2162-01-17 05:50:00	2162-01-17 05:50:00	EMERGENCY	CLINIC REFERRAL/PREMATURE	DEAD/EXPIRED	Pri

7 rows × 27 columns

```python
In [5]:  #Plot the age distribution of the Male patients that were admitted
         merged_data['age'][merged_data['GENDER'] == 'M'].describe()
```

```
Out[5]: count    31982.000000
        mean        52.873523
        std         25.893250
        min          0.000000
        25%         42.000000
        50%         59.000000
        75%         72.000000
        max         89.000000
        Name: age, dtype: float64
```

```python
In [6]:  #PLot the age distribution of the Female patients that were admitted
         merged_data['age'][merged_data['GENDER'] == 'F'].describe()
```

```
Out[6]: count    24378.000000
        mean        53.403889
        std         27.596749
        min          0.000000
        25%         40.000000
        50%         61.000000
        75%         75.000000
        max         89.000000
        Name: age, dtype: float64
```

```python
In [7]:  diagnoses = pd.read_csv('DIAGNOSES_ICD.csv')
         diagnosis_labels = pd.read_csv('D_ICD_DIAGNOSES.csv')
         icd9_dict = diagnosis_labels.set_index('ICD9_CODE')['SHORT_TITLE'].to_dict()
```

PUERTO RICAN DIAGNOSES BOX PLOT

In [8]:
```python
puerto_rican_admissions = merged_data[merged_data['ETHNICITY'] == 'HISPANIC/LATINO - PUERTO RICAN']

merged_puerto_ricans = pd.merge(puerto_rican_admissions, diagnoses, on=['HADM_ID','SUBJECT_ID'])

merged_puerto_ricans['SHORT_TITLE'] = merged_puerto_ricans['ICD9_CODE'].map(icd9_dict)
top_10_puerto_rican_diagnoses = merged_puerto_ricans['SHORT_TITLE'].value_counts().head(10)

top_puerto_rican_diagnoses = merged_puerto_ricans[merged_puerto_ricans['SHORT_TITLE'].isin(top_10_puerto_rican_diagnoses.index)]
top_puerto_rican_diagnoses['SHORT_TITLE'].value_counts()
```
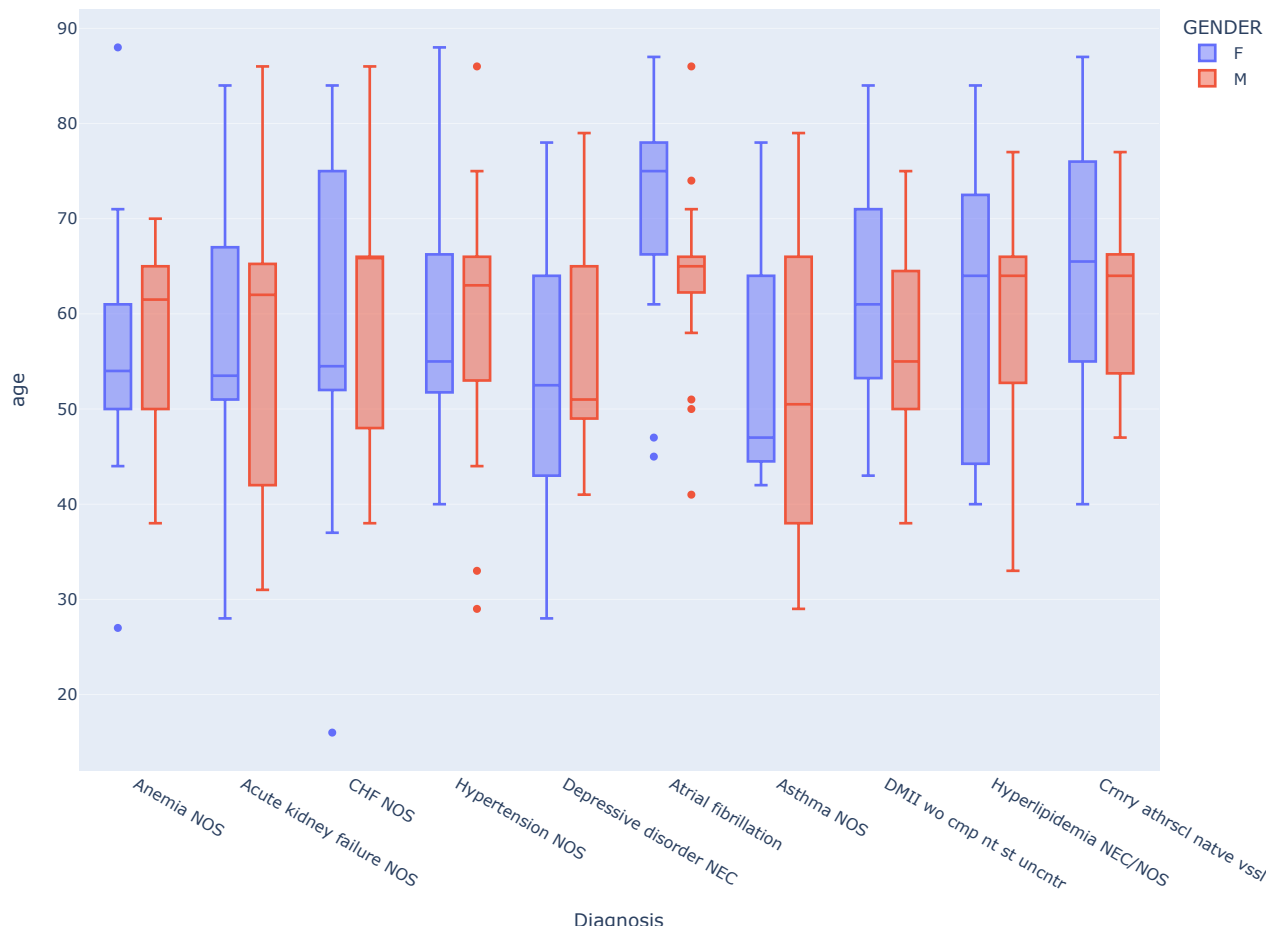
Out[8]:
```
Hypertension NOS          91
CHF NOS                   59
Hyperlipidemia NEC/NOS    56
Acute kidney failure NOS  55
DMII wo cmp nt st uncntr   48
Atrial fibrillation       46
Crnry athrscl natve vssl  43
Anemia NOS                36
Depressive disorder NEC   36
Asthma NOS                34
Name: SHORT_TITLE, dtype: int64
```

In [9]:
```python
fig = px.box(top_puerto_rican_diagnoses, x='SHORT_TITLE', y='age', color='GENDER', title='Top 10 Diagnoses for Puerto Rican Patients by Age and Gender',
             labels={'SHORT_TITLE':'Diagnosis'})
fig.update_layout(height=800)
fig.show()
```


Top 10 Diagnoses for Puerto Rican Patients by Age and Gender

AVERAGE ICUSTAY SUNBURST CHART

In [10]:
```python
icustays = pd.read_csv('icustays.csv')

merged_data = pd.merge(merged_data, icustays, on=['HADM_ID', 'SUBJECT_ID'])
```

In [11]:
```python
#Print the dataframe for one single patient to check to see the amount of admissions over the years
merged_data[merged_data['SUBJECT_ID'] == 188].head(10)
```

Out[11]:

	ROW_ID_x	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURAI
462	245	188	160697	2157-01-11 16:56:00	2157-01-19 14:58:00	NaN	EMERGENCY	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Pri
463	246	188	191517	2157-03-07 11:08:00	2157-03-10 13:50:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME	Pri
464	247	188	150463	2157-11-17 12:11:00	2157-11-20 13:05:00	NaN	EMERGENCY	CLINIC REFERRAL/PREMATURE	HOME	Pri
465	248	188	192557	2160-11-25 21:55:00	2160-11-28 12:42:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Pri
466	249	188	123860	2160-12-31 12:34:00	2161-01-08 19:50:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Pri
467	250	188	164735	2161-07-01 19:44:00	2161-07-10 20:00:00	NaN	EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Pri
468	251	188	132401	2161-11-01 17:48:00	2162-01-17 05:50:00	2162-01-17 05:50:00	EMERGENCY	CLINIC REFERRAL/PREMATURE	DEAD/EXPIRED	Pri
469	251	188	132401	2161-11-01 17:48:00	2162-01-17 05:50:00	2162-01-17 05:50:00	EMERGENCY	CLINIC REFERRAL/PREMATURE	DEAD/EXPIRED	Pri

8 rows × 37 columns

In [12]:
```python
top_icu_diagnoses = merged_data['DIAGNOSIS'].value_counts().head(25)
print(top_icu_diagnoses)
top_icu_diagnoses = merged_data[merged_data['DIAGNOSIS'].isin(top_icu_diagnoses.index)]
```

```
NEWBORN                                                    7840
PNEUMONIA                                                  1461
SEPSIS                                                     1197
CONGESTIVE HEART FAILURE                                    903
CORONARY ARTERY DISEASE                                     891
CHEST PAIN                                                  769
ALTERED MENTAL STATUS                                       716
INTRACRANIAL HEMORRHAGE                                     700
GASTROINTESTINAL BLEED                                      683
CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT /SDA   590
UPPER GI BLEED                                              581
ABDOMINAL PAIN                                              576
FEVER                                                       539
CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT/SDA    468
DIABETIC KETOACIDOSIS                                       464
SUBARACHNOID HEMORRHAGE                                     442
LOWER GI BLEED                                              424
S/P FALL                                                    417
HYPOTENSION                                                 411
STROKE;TELEMETRY;TRANSIENT ISCHEMIC ATTACK                  321
SEIZURE                                                     300
ACUTE RENAL FAILURE                                         280
PANCREATITIS                                                267
S/P MOTOR VEHICLE ACCIDENT                                  247
RESPIRATORY FAILURE                                         246
Name: DIAGNOSIS, dtype: int64
```

```
avg_los = top_icu_diagnoses.groupby(['GENDER', 'DIAGNOSIS'])['LOS'].mean().reset_index()

fig = px.sunburst(avg_los, path=['GENDER', 'DIAGNOSIS'], values='LOS', title='Average ICU LOS per Gender for the Top 25 Diagnoses')
fig.update_layout(width=1300, height=800)
fig.show()
```
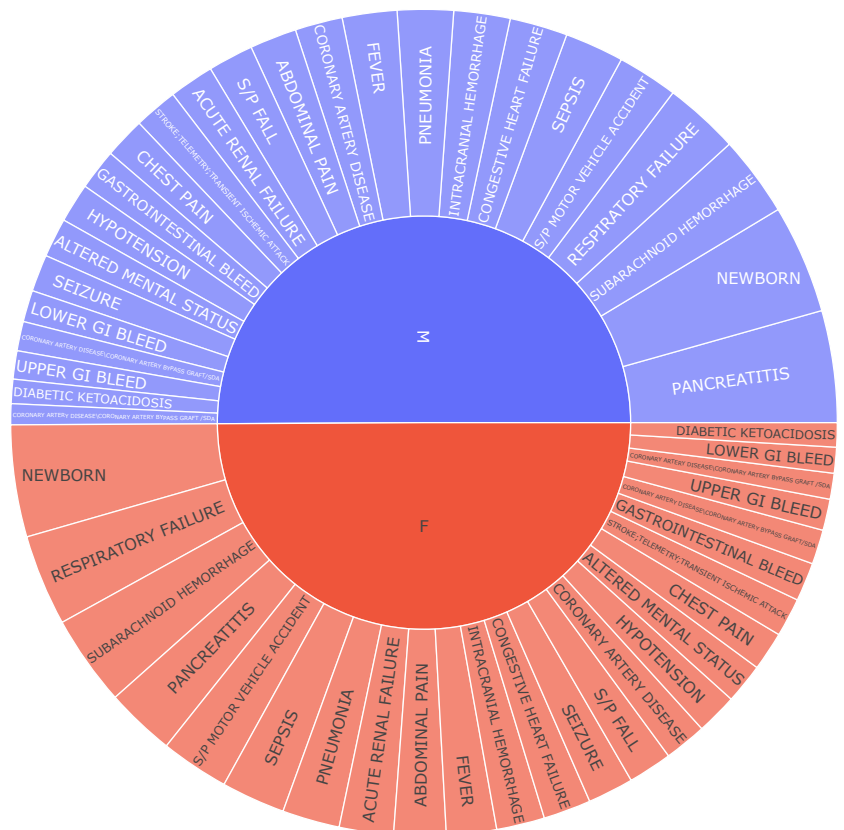
C:\Users\alejm\anaconda3\lib\site-packages\plotly\express_core.py:1637: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.

C:\Users\alejm\anaconda3\lib\site-packages\plotly\express_core.py:1637: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.

Average ICU LOS per Gender for the Top 25 Diagnoses

SUBARACHNOID HEMORRHAGE DENSITY HEATMAP

```python
# Search for the amount of subarachnoids in the dataset
subarachnoids = merged_data[merged_data['DIAGNOSIS'] == 'SUBARACHNOID HEMORRHAGE']

# Ensure that there is no duplicate data with the same patient and same diagnosis
subarachnoids = subarachnoids.drop_duplicates(subset=['HADM_ID', 'DIAGNOSIS'])
subarachnoids['MORTALITY'] = subarachnoids['DEATHTIME'].notnull()
subarachnoids['MORTALITY'] = subarachnoids['MORTALITY'].astype(str)
subarachnoids['MORTALITY'] = subarachnoids['MORTALITY'].replace({'True': 'DEATH', 'False': 'DISCHARGED'})

subarachnoids['MORTALITY'].value_counts()

#fig = px.histogram(subarachnoids, x='age', title='Distribution of Ages of Patient with subarachnoids and MORTALITY', color='M
ORTALITY', nbins=50)
#fig.show()

fig = px.density_heatmap(subarachnoids, x='age', y='MORTALITY', title='Density Heatmap of Age and Mortality for Subarachnoid H
emorrhages',
                         marginal_x='histogram', marginal_y='histogram', nbinsx=50, labels={'age':'Age', 'MORTALITY':'Mortalit
y'})
fig.update_layout(height=800)
fig.show()
```
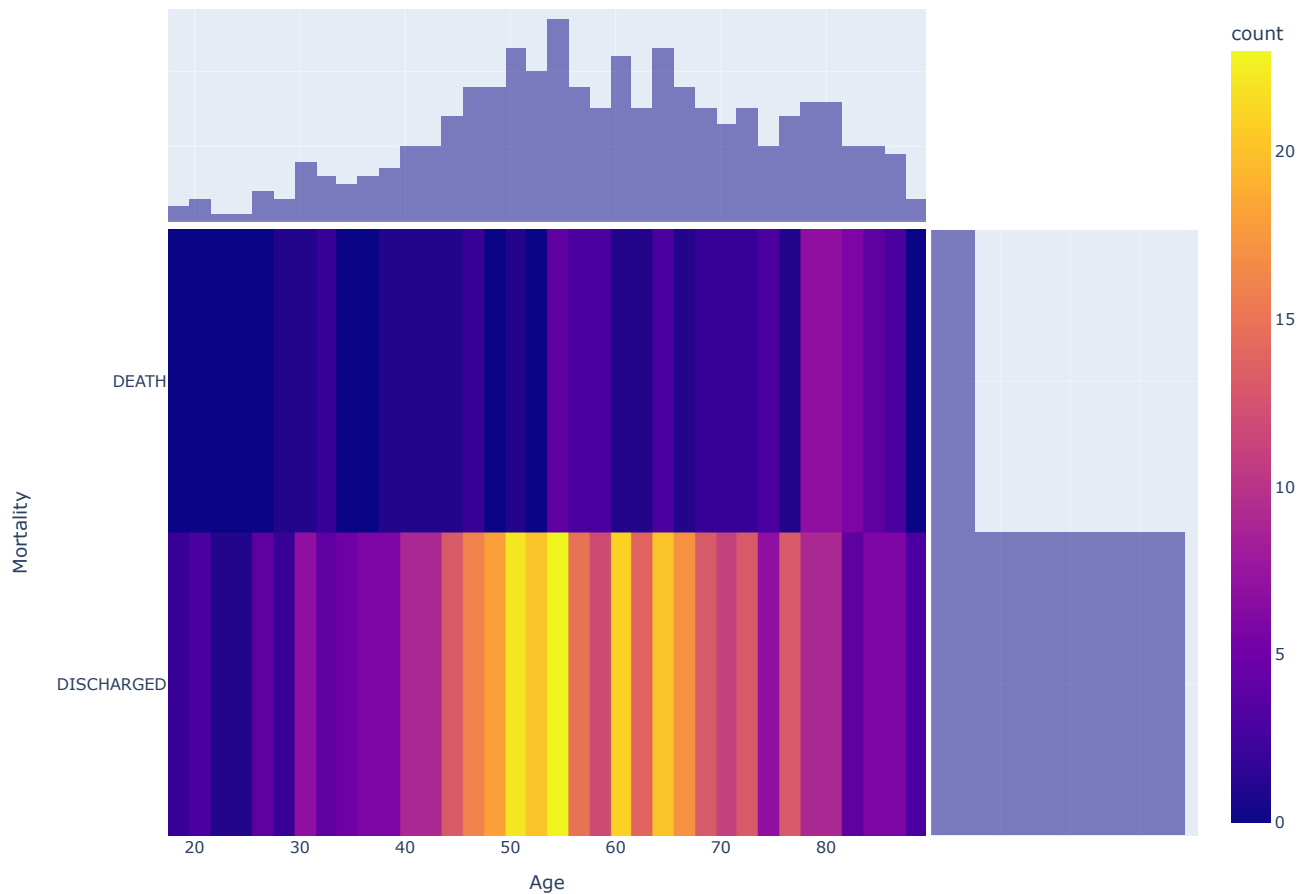
Density Heatmap of Age and Mortality for Subarachnoid Hemorrhages

TOP PROCDURES LINE PLOT

```python
procedures = pd.read_csv('PROCEDURES_ICD.csv')
procedure_labels = pd.read_csv('D_ICD_PROCEDURES.csv')
icd9_dict = procedure_labels.set_index('ICD9_CODE')['SHORT_TITLE'].to_dict()
```

```
In [16]: merged_procedures = pd.merge(admissions, procedures, on=['HADM_ID','SUBJECT_ID'])

         merged_procedures['SHORT_TITLE'] = merged_procedures['ICD9_CODE'].map(icd9_dict)
         top_10_procedures = merged_procedures['SHORT_TITLE'].value_counts().head(10)
         print(top_10_procedures)

         top_procedures = merged_procedures[merged_procedures['SHORT_TITLE'].isin(top_10_procedures.index)]
```

```
Venous cath NEC            14731
Insert endotracheal tube   10333
Entral infus nutrit sub     9300
Cont inv mec ven <96 hrs    9100
Packed cell transfusion     7244
Extracorporeal circulat     6838
Cont inv mec ven 96+ hrs    6048
Vaccination NEC             5842
Coronar arteriogr-2 cath    5337
Arterial catheterization    4737
Name: SHORT_TITLE, dtype: int64
```

```python
#extract year from the admission date
top_procedures['ADMITTIME'] = pd.to_datetime(top_procedures['ADMITTIME'])
top_procedures['YEAR'] = top_procedures['ADMITTIME'].dt.year

procedure_stats = top_procedures.groupby('YEAR')['SHORT_TITLE'].value_counts().unstack().fillna(0)

fig = px.line(procedure_stats, title='Top 10 All-Time Procedures Breakdown by Year', labels={'value':'Number of Procedures',
'YEAR':'Year', 'SHORT_TITLE':'Procedure'})
fig.update_layout(height=800)
fig.show()
```

C:\Users\alejm\AppData\Local\Temp\ipykernel_11440\3065196708.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
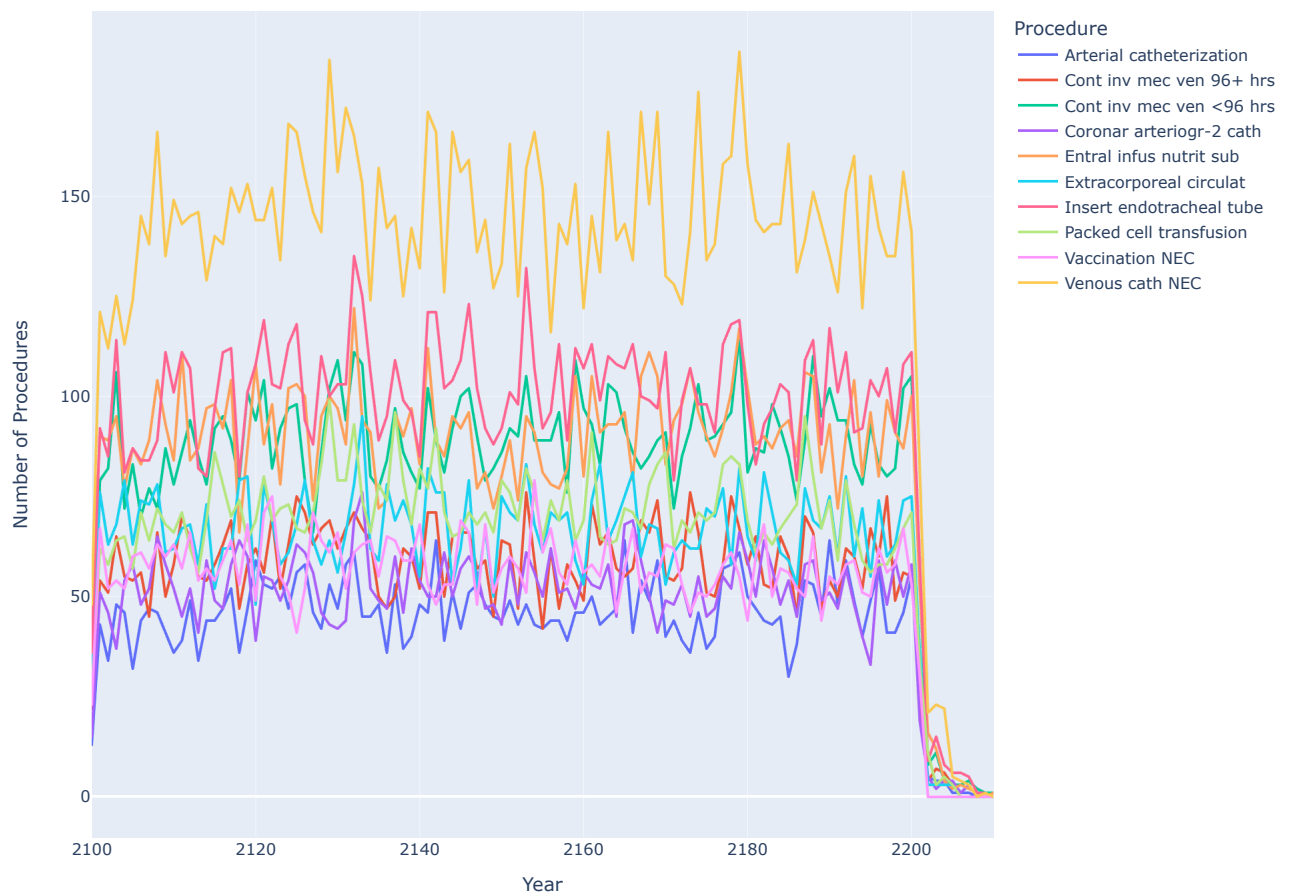Try using .loc[row_indexer,col_indexer] = value instead

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

C:\Users\alejm\AppData\Local\Temp\ipykernel_11440\3065196708.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

Top 10 All-Time Procedures Breakdown by Year

GENDER, MARITAL STATUS, DIAGNOSIS PARRALEL CATEGORIES DIAGRAM

```
In [18]: top_marital_demos = merged_data['DIAGNOSIS'][merged_data['MARITAL_STATUS'].isin(['MARRIED', 'SINGLE', 'DIVORCED', 'WIDOWE
         D'])].value_counts().head(10)
         print(top_marital_demos)
         top_marital_demos = merged_data[merged_data['DIAGNOSIS'].isin(top_marital_demos.index)]
         top_marital_demos = top_marital_demos[top_marital_demos['MARITAL_STATUS'].isin(['MARRIED', 'SINGLE', 'DIVORCED', 'WIDOWED'])]

         fig = px.parallel_categories(top_marital_demos, dimensions=['GENDER', 'MARITAL_STATUS', 'DIAGNOSIS'], title='Parallel Categori
         es of Gender, Marital Status, and Diagnosis')
         fig.update_layout(height=800)
         fig.show()
```

```
PNEUMONIA                                              1395
SEPSIS                                                 1122
CONGESTIVE HEART FAILURE                                863
CORONARY ARTERY DISEASE                                 848
CHEST PAIN                                              739
GASTROINTESTINAL BLEED                                  657
ALTERED MENTAL STATUS                                   646
INTRACRANIAL HEMORRHAGE                                 612
CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT /SDA    579
ABDOMINAL PAIN                                          547
Name: DIAGNOSIS, dtype: int64
```


Parallel Categories of Gender, Marital Status, and Diagnosis