Analyzing Healthcare Trends

Group 1 Team Members: Brianna Bethea,Kavita Gopal, Laquita Palmer, Derek Hill

Project Proposal: Introduction

- Introduction- Understanding health trends is essential for identifying emerging health issues, shaping public health policies, and enhancing healthcare services. This project aims to analyze key health trends by examining prevalent medical conditions, lifestyle factors, and demographic influences. The objective is to uncover patterns and correlations that can provide actionable insights for stakeholders.Our team, consisting of Brianna, Derek, Kavita, and Laquita, has conducted an in-depth analysis using the healthcare dataset from the "Unlocking Healthcare Trends: Data Analysis" project authored by Muhammad Furqan. The dataset, sourced from Kaggle, contains 10,000 synthetic patient records, which include a wide range of attributes such as patient demographics, medical conditions, and admission details. The data is entirely synthetic and is intended for educational and non-commercial use.The analysis was conducted using Python, along with libraries such as Pandas and Matplotlib, to perform data cleaning, analysis, and visualizations. These tools enabled us to explore the relationships among various variables in the dataset and draw meaningful conclusions.

Project Proposal: Data Sources & Objectives

Primary Dataset: Kaggle Health Dataset (health_dataset.csv)

The objectives of this project are as follows:

1. **Analyze Gender-Based Susceptibility to Medical Conditions:**
    - Identify the prevalence of medical conditions among male and female populations within the dataset.
    - Determine which gender is more susceptible to specific medical conditions.
2. **Evaluate Billing Amounts for One-Day Hospital Stays:**
    - Use statistical analysis to determine the highest billing amounts for various medical conditions during a one-day hospital stay.
3. **Examine Insurance Provider Coverage:**
    - Identify the most prevalent medical conditions covered by different insurance providers.
    - Compare the average billing amounts associated with each medical condition by insurance provider.
4. **Analyze Hospital Admission Types by Age Group:**
    - Compare the frequency of different hospital admission types (Elective, Emergency, Urgent) across various age groups.
5. **Determine the Most Commonly Admitted Conditions:**
    - Identify the medical condition with the highest admission rates and the corresponding insurance provider with the highest admission frequency.

Project Proposal: Methodology

Methodology

1. **Data Collection:**
   - Extract relevant data from the dataset, focusing on health conditions, billing information, insurance providers, duration of hospital stays, and admission types.
2. **Data Cleaning and Preparation:**
   - Clean and wrangle the data to address missing values, outliers, and inconsistencies.
   - Merge datasets as needed to create a comprehensive dataset for analysis.
3. **Exploratory Data Analysis (EDA):**
   - Conduct EDA to understand data distribution and identify patterns or trends.
   - Generate summary statistics for each variable to provide a foundational understanding of the dataset.
4. **Visualization:**
   - Create various visualizations, such as bar charts, heatmaps, pie charts, and bell curves, to illustrate the trends and correlations in the data.
   - Utilize these visualizations to present findings in a clear and comprehensible manner.
5. **Interpretation and Recommendations:**
   - Interpret the results of the analysis to draw conclusions about the health trends identified.
   - Provide recommendations based on these findings to inform healthcare policies and practices.

Project Proposal: Conclusions

1. **Insurance Provider and Medical Condition Correlation:**
   - **Conclusion:** Arthritis and Asthma are most commonly associated with Cigna, Cancer and Hypertension with United Healthcare, Diabetes with Medicare, and Obesity with Blue Cross.
2. **Admission Types by Generation:**
   - **Conclusion:** Generation X has the highest rates of Elective and Urgent admission types, while Seniors exhibit the highest rates of Emergency admissions.
3. **Billing Amounts for One-Day Hospital Stays:**
   - **Conclusion:** Hypertension incurs the highest average billing cost for a one-day hospital stay, with significant variance in billing amounts as depicted by a bell curve.
4. **Gender-Based Prevalence of Medical Conditions:**
   - **Conclusion:** Females show a higher prevalence of Arthritis, Diabetes, and Hypertension, with Obesity reaching 100%. Males are more prone to Asthma, Cancer, Diabetes, Arthritis, and Hypertension. Hypertension is equally distributed among both genders.
5. **Trends in Admissions Over Time:**
   - **Conclusion:** There was a steady increase in admissions into 2020, followed by a decline to regular admission numbers later. Additionally, older age is correlated with a higher prevalence of health conditions.

Team Member: Brianna Bethea

Focus: Medical Conditions/Insurance Providers; Admission Type/ Age

Objective:

1. To identify the most prevalent medical conditions covered by different insurance providers.

2. To compare the frequency of different hospital admission types (Elective, Emergency, Urgent) among various age groups.

Conclusions:

1. Arthritis & Asthma are most associated with Cigna, Cancer & Hypertension with United Healthcare, Diabetes with Medicare, and Obesity with Blue Cross.

2. Gen X has the highest Elective and Urgent admission types, while Seniors have the highest Emergency admission type.

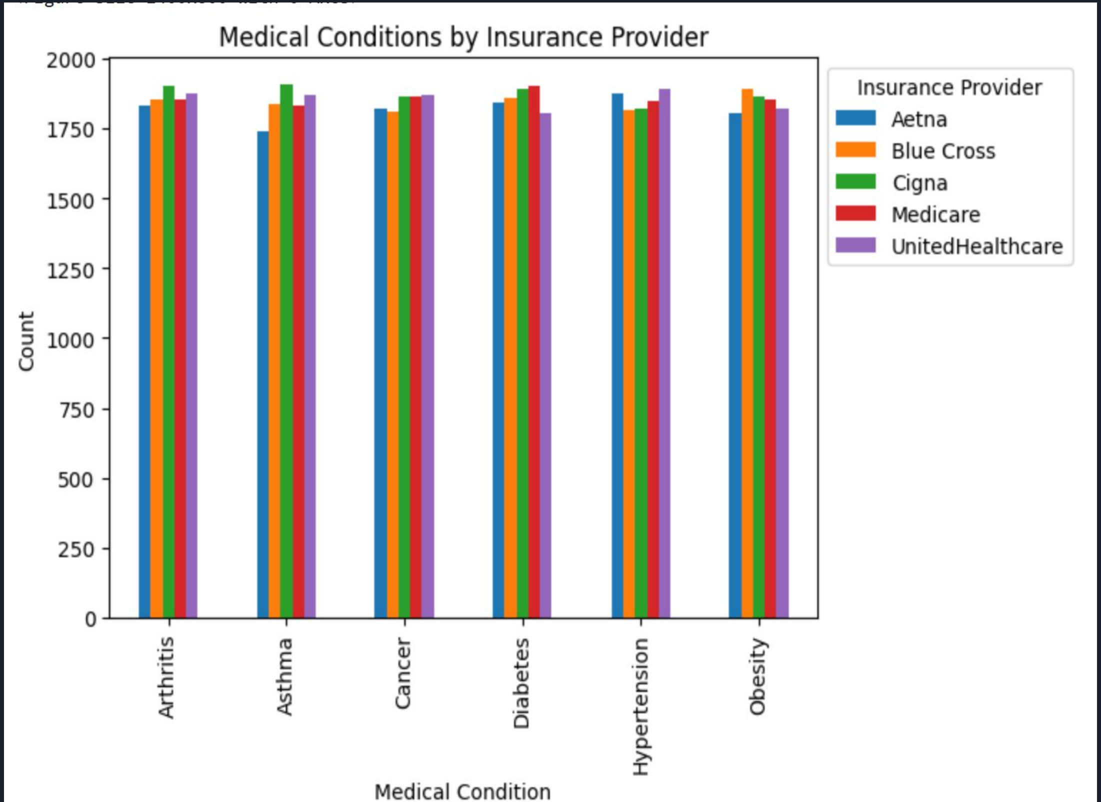
Groupby Dataset: Medical Condition & Insurance Provider

[203]:
```
# Show groupby data for medical condition and insurance provider
plt.figure(figsize=(14, 8))
health_data.groupby(['Medical Condition', 'Insurance Provider']).size().unstack()
```

[203]:

Insurance Provider	Aetna	Blue Cross	Cigna	Medicare	UnitedHealthcare
Medical Condition					
Arthritis	1832	1852	1900	1851	1873
Asthma	1740	1835	1907	1833	1870
Cancer	1819	1808	1864	1866	1870
Diabetes	1842	1860	1893	1903	1806
Hypertension	1876	1813	1821	1847	1888
Obesity	1804	1891	1864	1854	1818

Chart 1: Medical Conditions by Insurance Providers



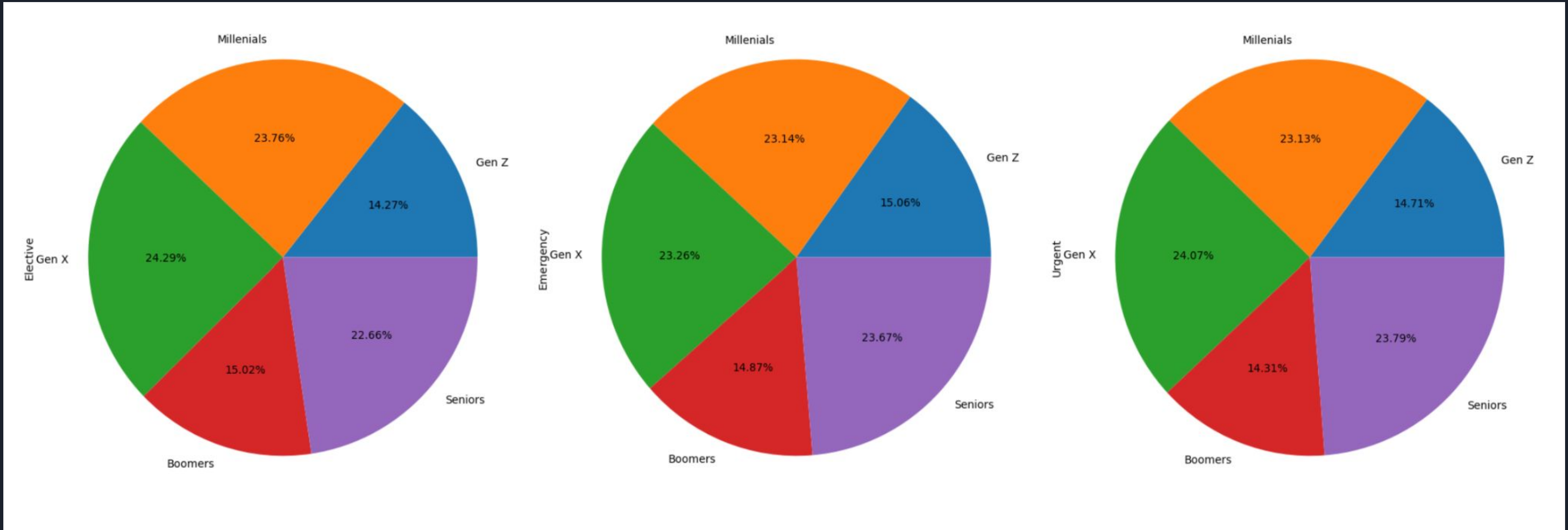
[144]:
```python
#Group data by Age and count the type of admissions by age

#grouped_health = health_data.groupby(["Age","Admission Type"]).count().reset_index()
grouped_health = health_data.groupby(['Age','Admission Type']).size().fillna(0)
grouped_health
```

[144]:
```
Age  Admission Type
13   Elective           2
     Emergency          8
     Urgent             4
14   Elective           8
     Emergency          7
                       ..
88   Emergency          6
     Urgent            10
89   Elective           5
     Emergency          2
     Urgent             1
Length: 231, dtype: int64
```

[165]:
```python
#Create bins for age ranges
bins = [12, 27, 43, 59, 69, 100]

#Create labels for each age range
group_ages = ['Gen Z', 'Millenials', 'Gen X', 'Boomers', 'Seniors']
print(group_ages)
```

```
['Gen Z', 'Millenials', 'Gen X', 'Boomers', 'Seniors']
```

Chart 2: Hospital Admission Type Across Generations



Kavita Gopal

- ★ Find the contents of the dataset
- ★ Cleaning up the dataset, to get a clearer idea, to enable suitable analysis & conclusion
- ★ Aim was to combine cost of hospitalization-duration, with the treatment of the specific medical condition
- ★ Aim to get some data on gender-prone to diseases within the dataset.

healthcare_df

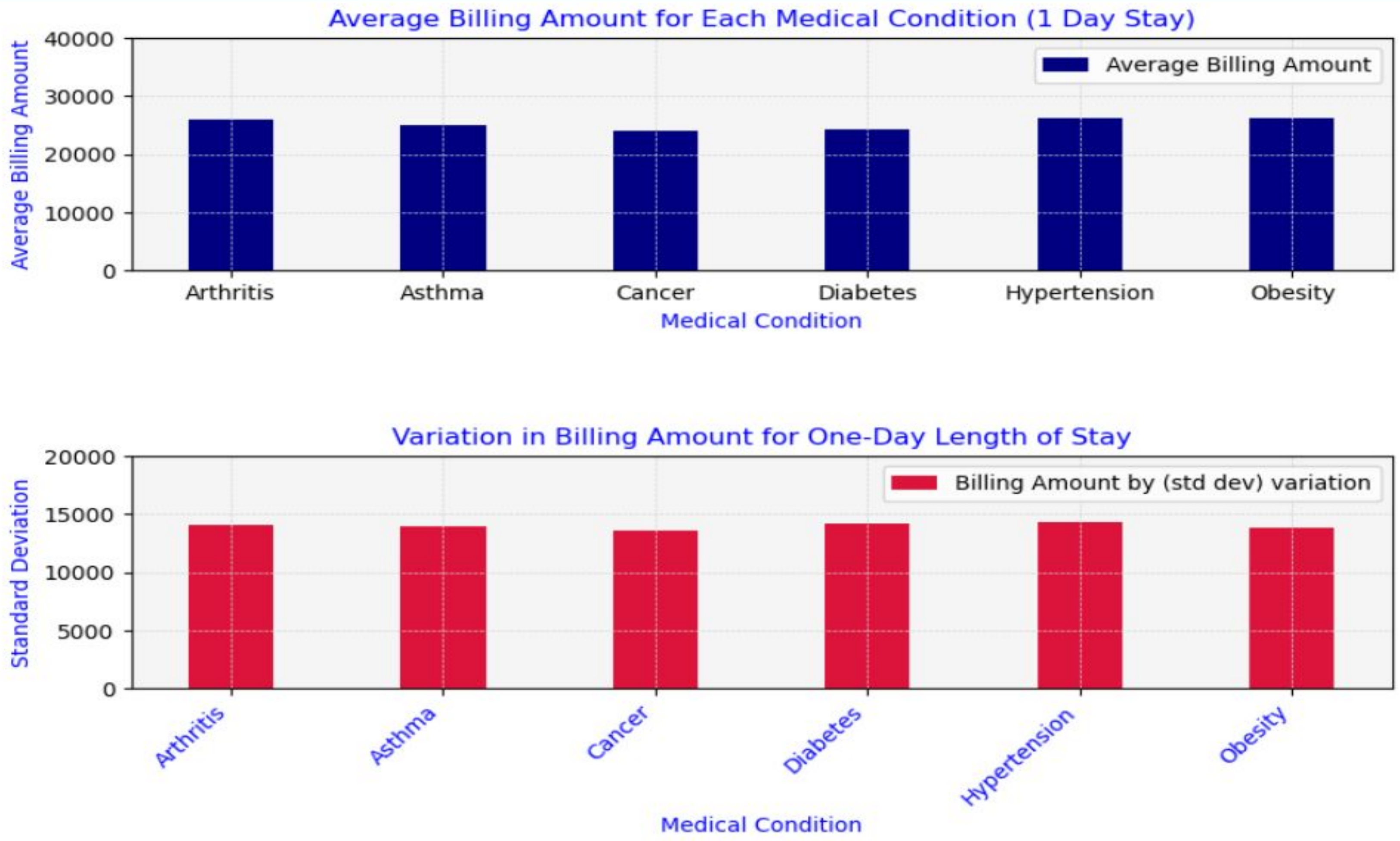
	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication
0	Bobby JacksOn	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18856.281306	328	Urgent	2024-02-02	Paracetamol
1	LesLie TErRy	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	265	Emergency	2019-08-26	Ibuprofen
2	DaNnY sMitH	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096079	205	Emergency	2022-10-07	Aspirin
3	andrEw waTtS	28	Female	O+	Diabetes	2020-11-18	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.782410	450	Elective	2020-12-18	Ibuprofen
4	adrIENNE bEll	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Aetna	14238.317814	458	Urgent	2022-10-09	Penicillin
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	eLIZABeTH jaCkSOn	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2650.714952	417	Elective	2020-09-15	Penicillin
55496	KYle pEREz	61	Female	AB-	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	316	Elective	2020-02-01	Aspirin
55497	HEATher WaNG	38	Female	B+	Hypertension	2020-07-13	Joe Jacobs DVM	and Mahoney Johnson Vasquez,	UnitedHealthcare	27620.764717	347	Urgent	2020-08-10	Ibuprofen
55498	JENniFER JOneS	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro,	Medicare	32451.092358	321	Elective	2019-05-31	Ibuprofen
55499	jAMES GARCiA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.134172	448	Urgent	2024-04-29	Ibuprofen

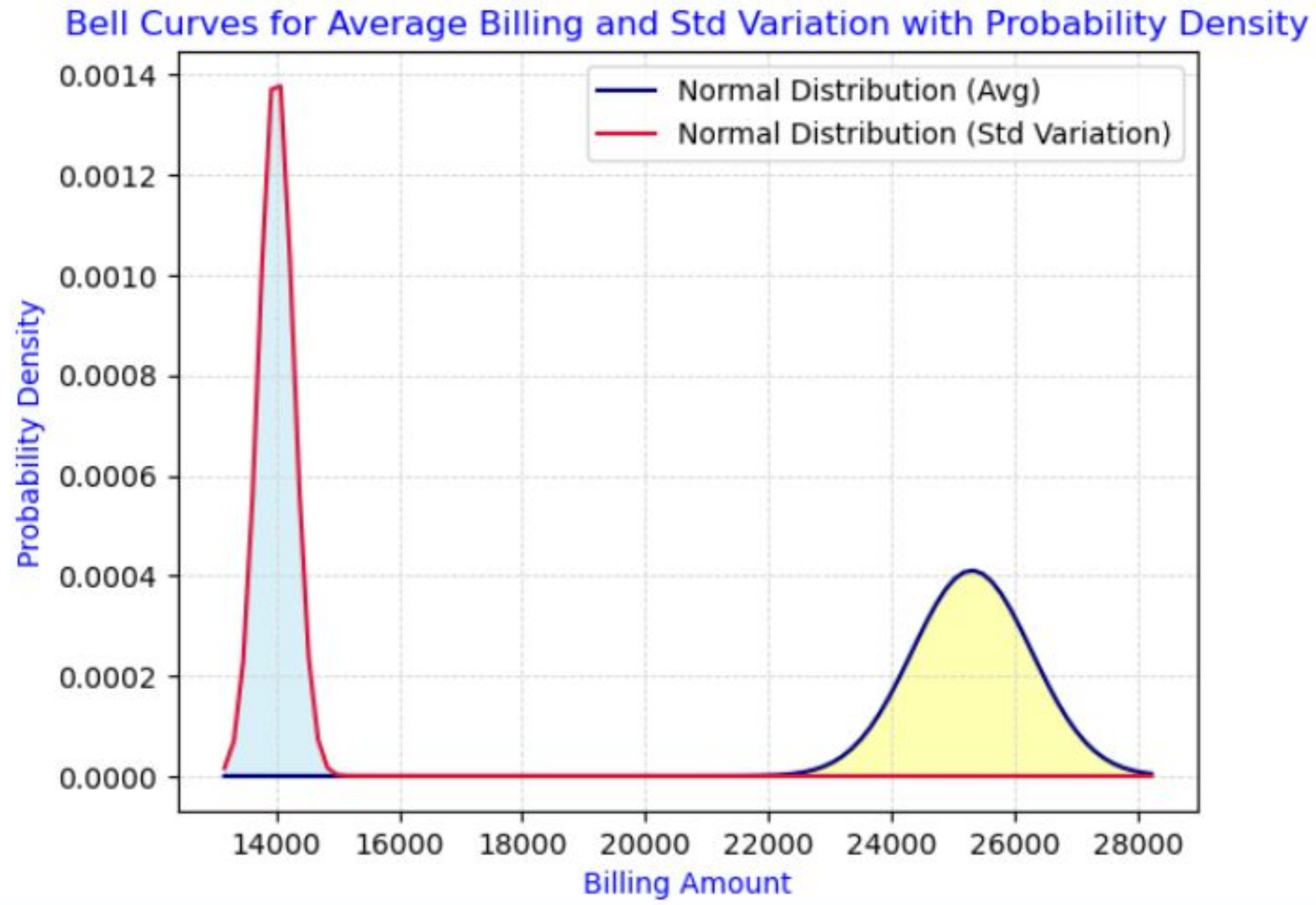
55500 rows × 15 columns

```python
# Getting the mean of one day length of stay by grouping medical condition and billing amount
avg_bill_amt_1d = one_day_length_of_stay_df.groupby("Medical Condition")["Billing Amount"].mean().round(2).reset_index()
avg_bill_amt_1d
```

	Medical Condition	Billing Amount
0	Arthritis	26016.13
1	Asthma	25088.88
2	Cancer	24039.15
3	Diabetes	24319.64
4	Hypertension	26215.13
5	Obesity	26182.85

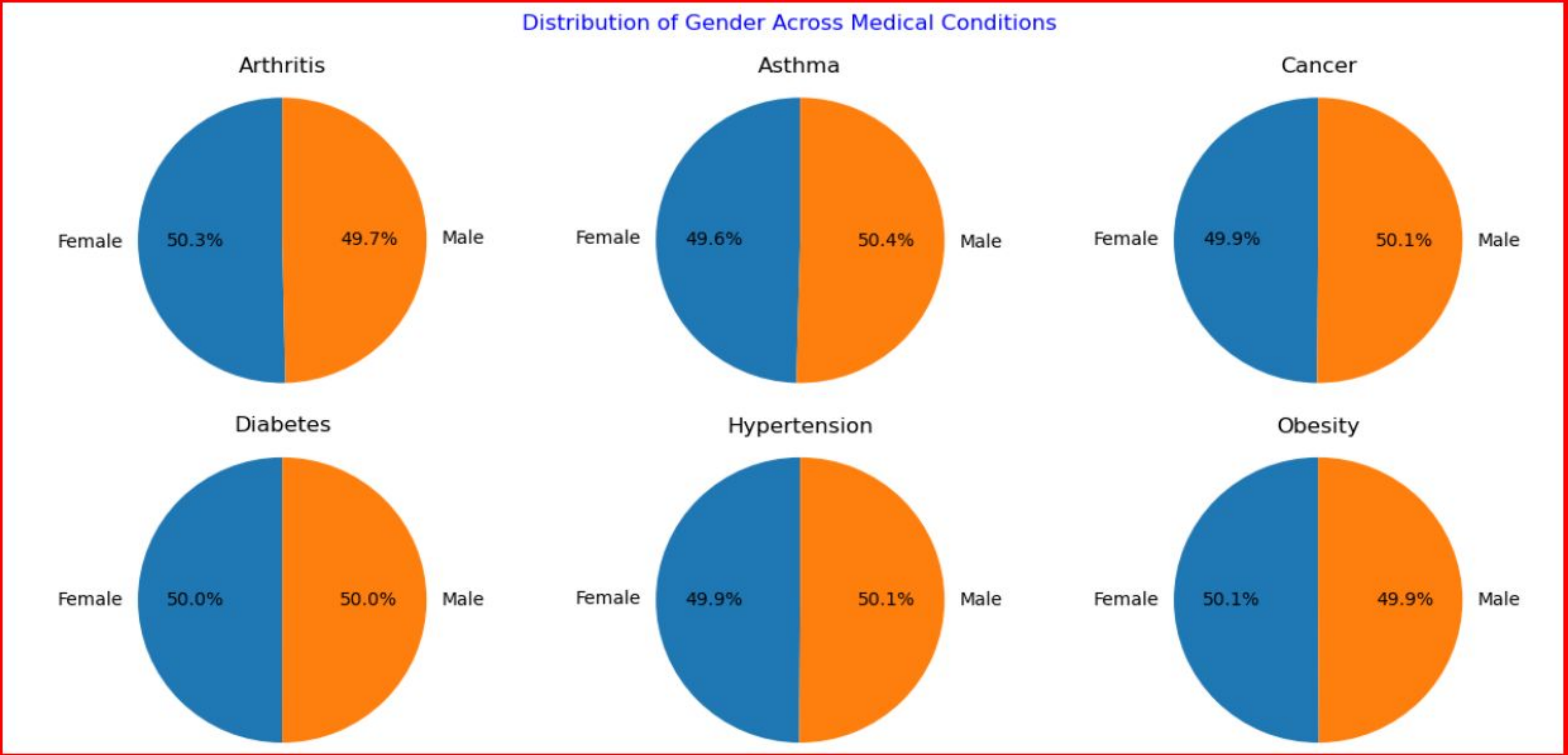
	Medical Condition	Billing Amount
0	Arthritis	14103.13
1	Asthma	13941.31
2	Cancer	13574.72
3	Diabetes	14171.87
4	Hypertension	14374.70
5	Obesity	13837.04


Average Billing Amount for Each Medical Condition (1 Day Stay)
Average Billing Amount
40000
30000
20000
10000
0
Average Billing Amount
Arthritis
Asthma
Cancer
Diabetes
Hypertension
Obesity
Medical Condition
Variation in Billing Amount for One-Day Length of Stay
Standard Deviation
20000
15000
10000
5000
0
Billing Amount by (std dev) variation
Arthritis
Asthma
Cancer
Diabetes
Hypertension
Obesity
Medical Condition


Bell Curves for Average Billing and Std Variation with Probability Density
Normal Distribution (Avg)
Normal Distribution (Std Variation)
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
Probability Density
14000
16000
18000
20000
22000
24000
26000
28000
Billing Amount

```python
# Display values of the most common medical condition for each gender:
gender_condition_counts = healthcare_df.groupby(['Gender', 'Medical Condition'])['Medical Condition'].count()
print(gender_condition_counts)
```

```
Gender  Medical Condition
Female  Arthritis            4686
        Asthma               4553
        Cancer               4602
        Diabetes             4651
        Hypertension         4612
        Obesity              4622
Male    Arthritis            4622
        Asthma               4632
        Cancer               4625
        Diabetes             4653
        Hypertension         4633
        Obesity              4609
```


Distribution of Gender Across Medical Conditions
Arthritis
Female
50.3%
49.7%
Male
Asthma
Female
49.6%
50.4%
Male
Cancer
Female
49.9%
50.1%
Male
Diabetes
Female
50.0%
50.0%
Male
Hypertension
Female
49.9%
50.1%
Male
Obesity
Female
50.1%
49.9%
Male

Conclusions:

- One-day hospital stay for managing hypertension had the highest billing cost in this data-set.
- The higher probability density peak and wider spread in standard deviation plot suggested more variability in terms of individual billing amounts. Also, indicative of potential outliers, (high/low billing amounts) contributing to the variability
- The prevalence of medical conditions amongst the females was slightly more for arthritis & obesity. Males were slightly more prone for asthma, cancer & hypertension. Both Males & females were equally prone for diabetes.

Thank you: Team members, Richard, & Deborah, and all of you, for your Support!

Team Member: LaQuita Palmer

Data Origin: We have analyzed data from the csv entitled Unlocking Healthcare Trends, authored by Muhammad Furqan.

Focus: In this dataset, I have grouped the data according to 2 different criteria: By Insurance Provider, Medical Condition, and Billing Amount, and by Insurance Provider and the number of admissions for each Provider.

Visualizations: From these groupbys I have visualized the following: 1) Average Billing Amount by Insurance Provider and Medical Condition 2) Most admitted Medical Condition.

I used the following visualizations: Bar Chart with Subplots and Pie Chart.

Conclusions:

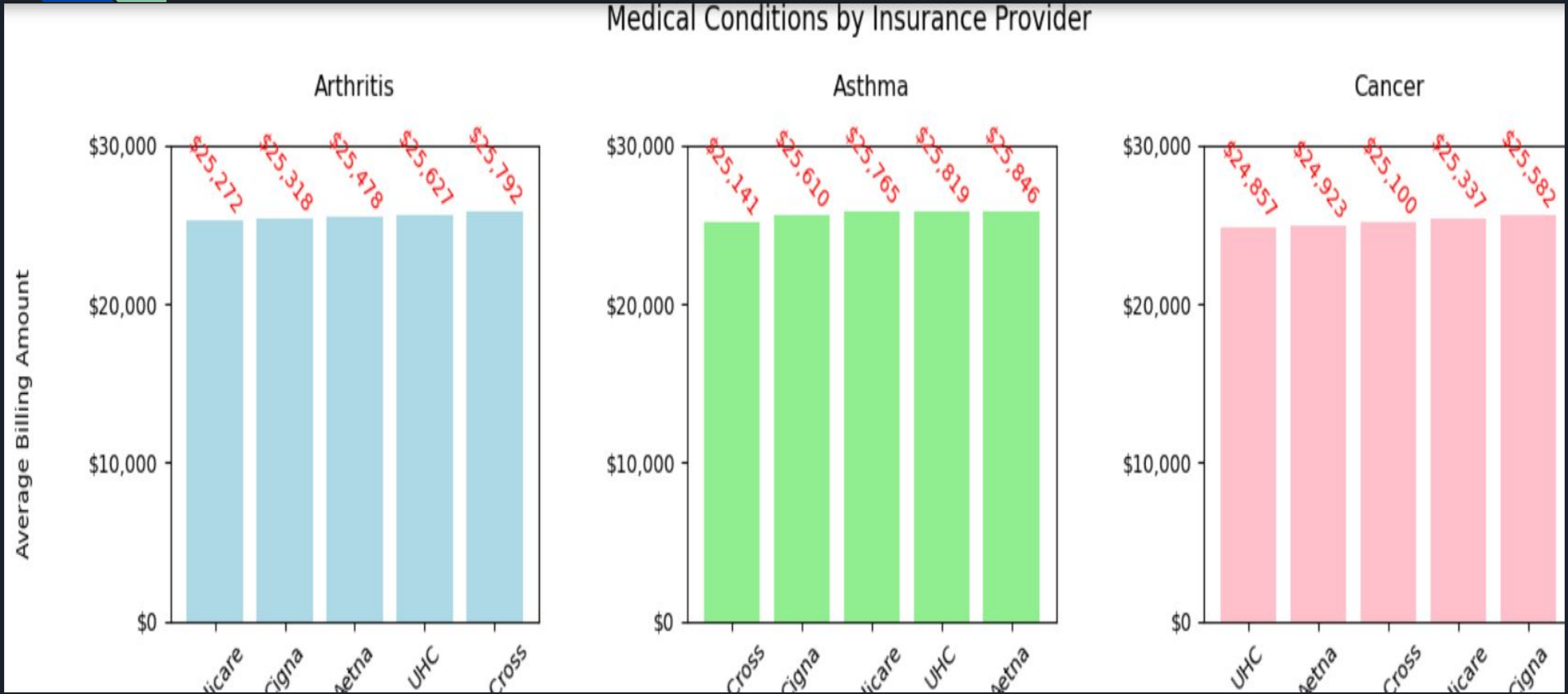
Conclusion1: The highest Provider and Billing Amount: BlueCross @ $26100.78 and Cigna @ $26116.99 (Obesity)

Conclusion2:: The most admitted condition: Arthritis and Diabetes both at 16.8%

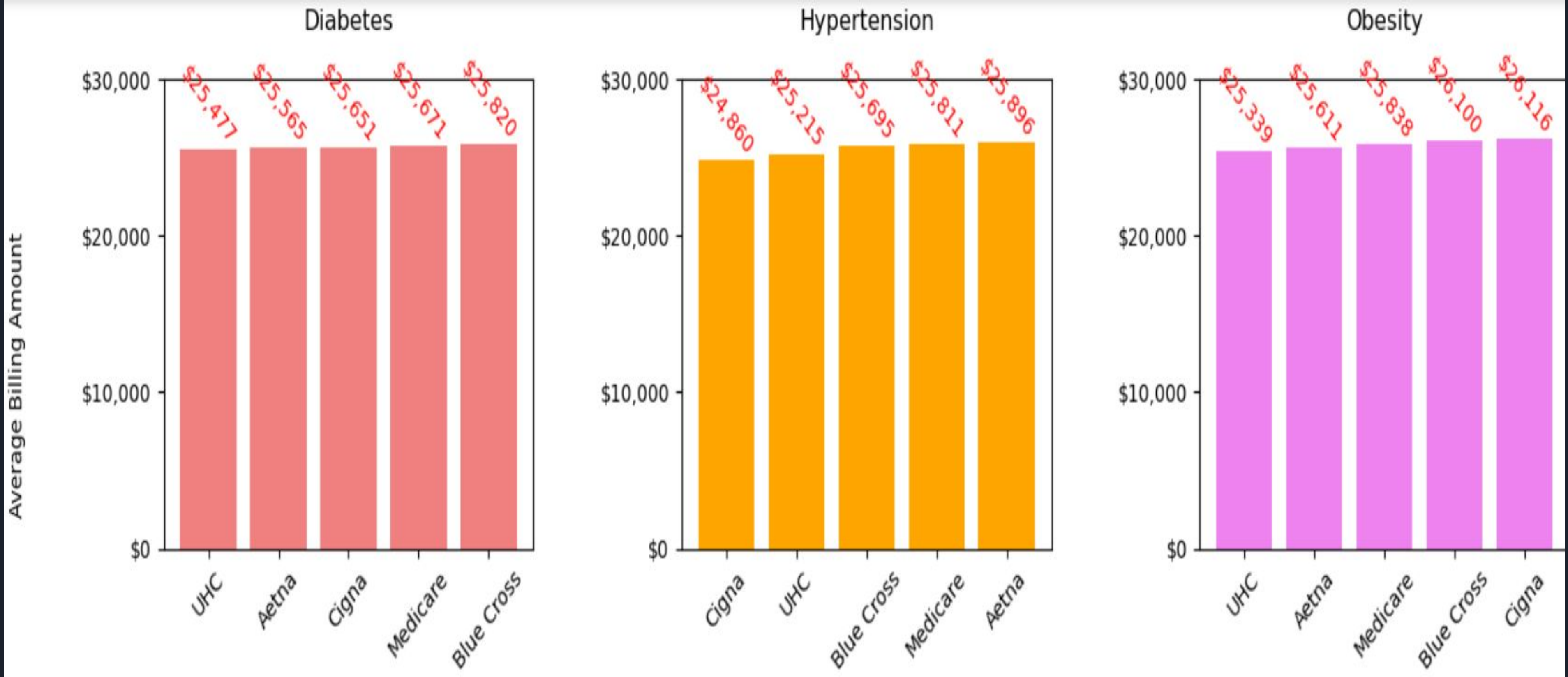
```python
# Group the data by Insurance Provider, Medical Condition, and Billing Amount
ins_conditions = data.groupby(['Insurance Provider', 'Medical Condition'])['Billing Amount'].mean().unstack()
new_ins_conditions=ins_conditions.reset_index()
new_ins_conditions
```

```python
#Melt the grouped data
ins_conditions_melted = new_ins_conditions.melt(id_vars=['Insurance Provider'],
                    var_name='Medical Condition',
                    value_name='Value')
ins_conditions_melted
insurance = ins_conditions_melted['Insurance Provider'].unique()

conditions = ins_conditions_melted['Medical Condition'].unique()

# Replace the names
ins_conditions_melted['Insurance Provider'] = ins_conditions_melted['Insurance Provider'].replace({
    'UnitedHealthcare': 'UHC'})
ins_conditions_melted
```

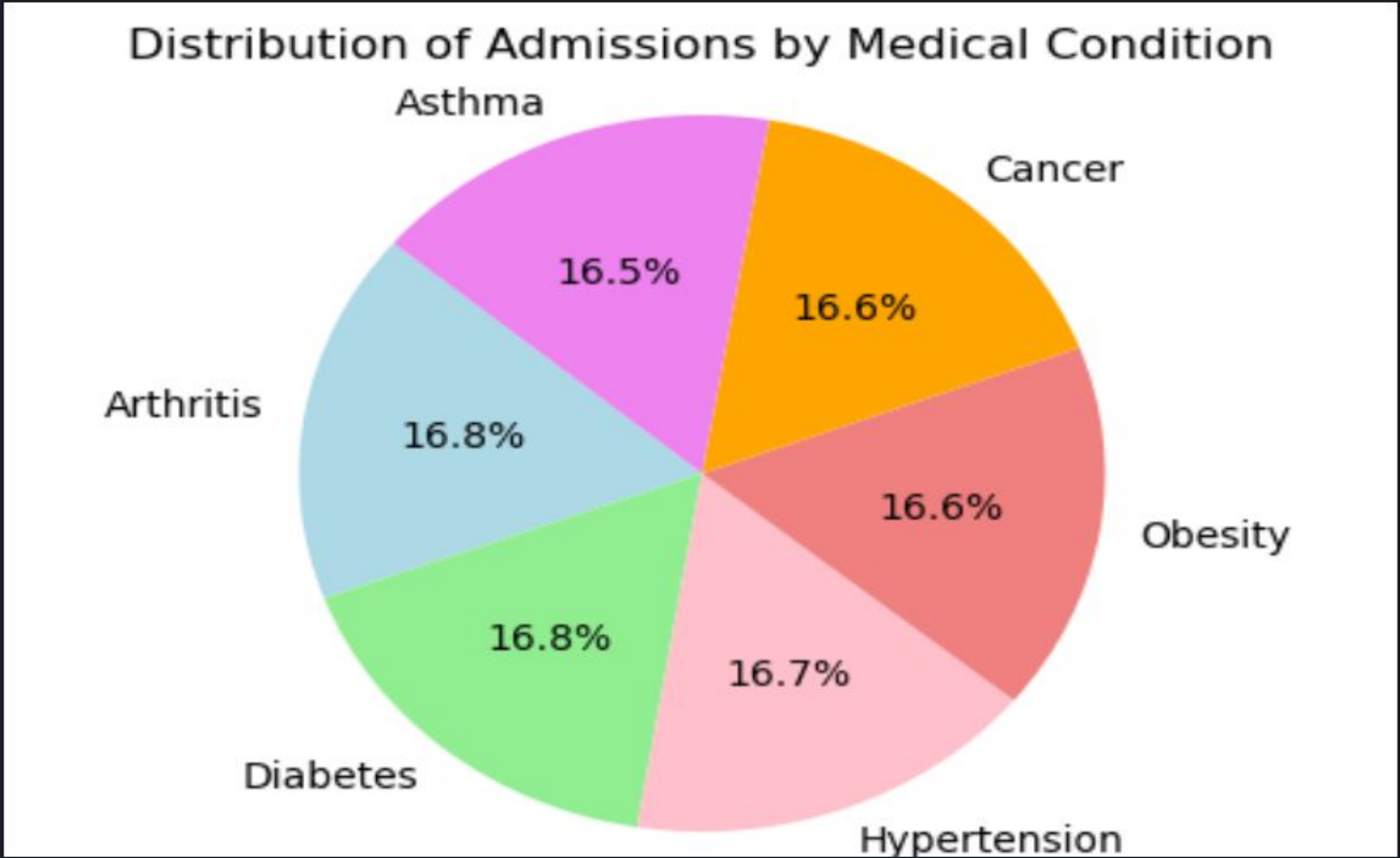
```python
# Identify the Medical Condition that patients are being admitted the most for
most_admitted_conditions = data.groupby('Medical Condition')['Date of Admission'].count().reset_index()
most_admitted_conditions = most_admitted_conditions.sort_values(by='Date of Admission', ascending=False)
most_admitted_conditions
```



Medical Conditions by Insurance Provider
Arthritis
Asthma
Cancer
Average Billing Amount
$30,000
$20,000
$10,000
$0
$25,272
$25,318
$25,478
$25,627
$25,792
$25,141
$25,610
$25,765
$25,819
$25,846
$24,857
$24,923
$25,100
$25,337
$25,582
Cigna
Aetna
UHC
Cross
Cigna
UHC
Aetna
UHC
Aetna
Cross
Cigna



Diabetes
Hypertension
Obesity
Average Billing Amount
$30,000
$20,000
$10,000
$0
$25,477
$25,565
$25,651
$25,671
$25,820
UHC
Aetna
Cigna
Medicare
Blue Cross
$24,860
$25,215
$25,695
$25,811
$25,896
Cigna
UHC
Blue Cross
Medicare
Aetna
$25,339
$25,611
$25,838
$26,100
$26,116
UHC
Aetna
Medicare
Blue Cross
Cigna


Distribution of Admissions by Medical Condition
Asthma
16.5%
Cancer
16.6%
Arthritis
16.8%
Obesity
16.6%
Diabetes
16.8%
Hypertension
16.7%

Derek Hill

Questions:

1. What are the admission trends over time?

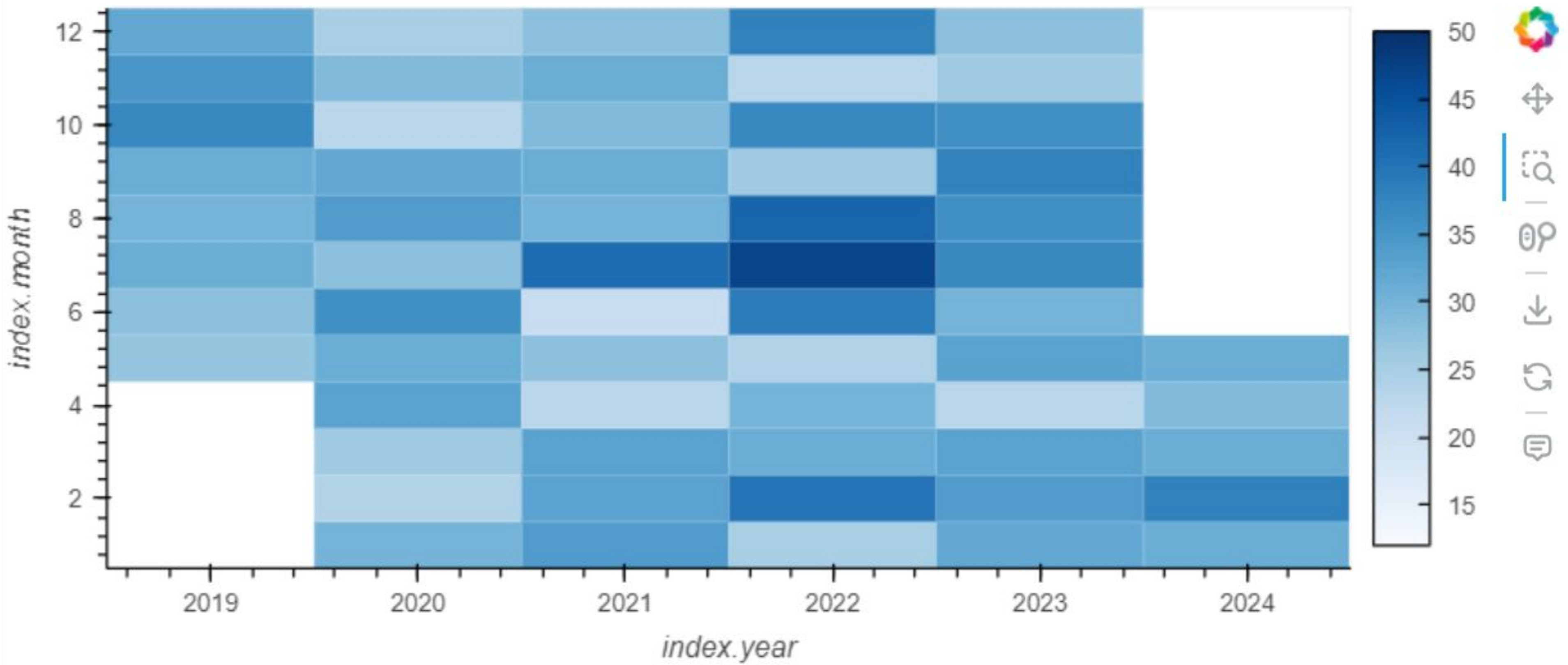
2. What is the age distribution among medical conditions?

Conclusions:

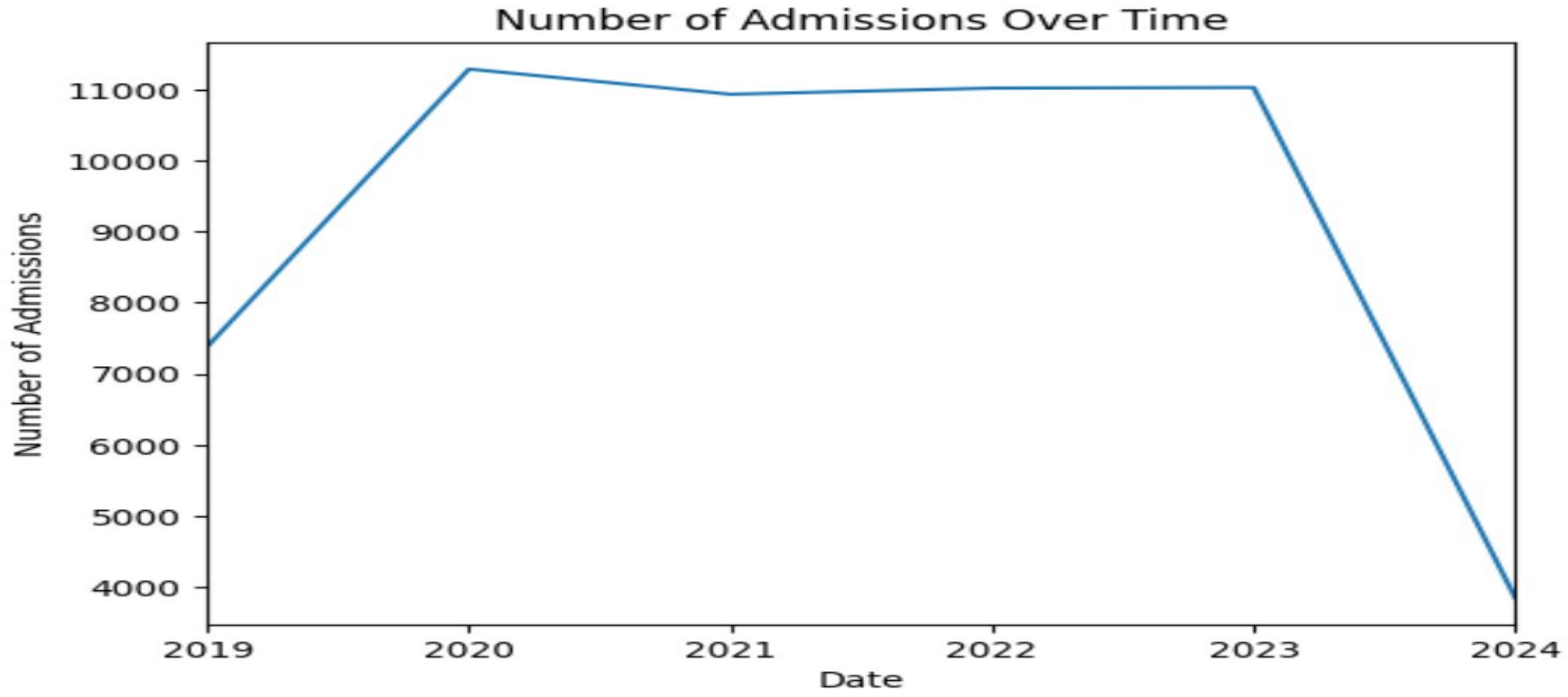
1. There is a steady increase into 2020, but declines to regular admission numbers later on.

2. There is obvious spikes of conditions among patients 20 and older(Older age is correlated with decreased health usually)

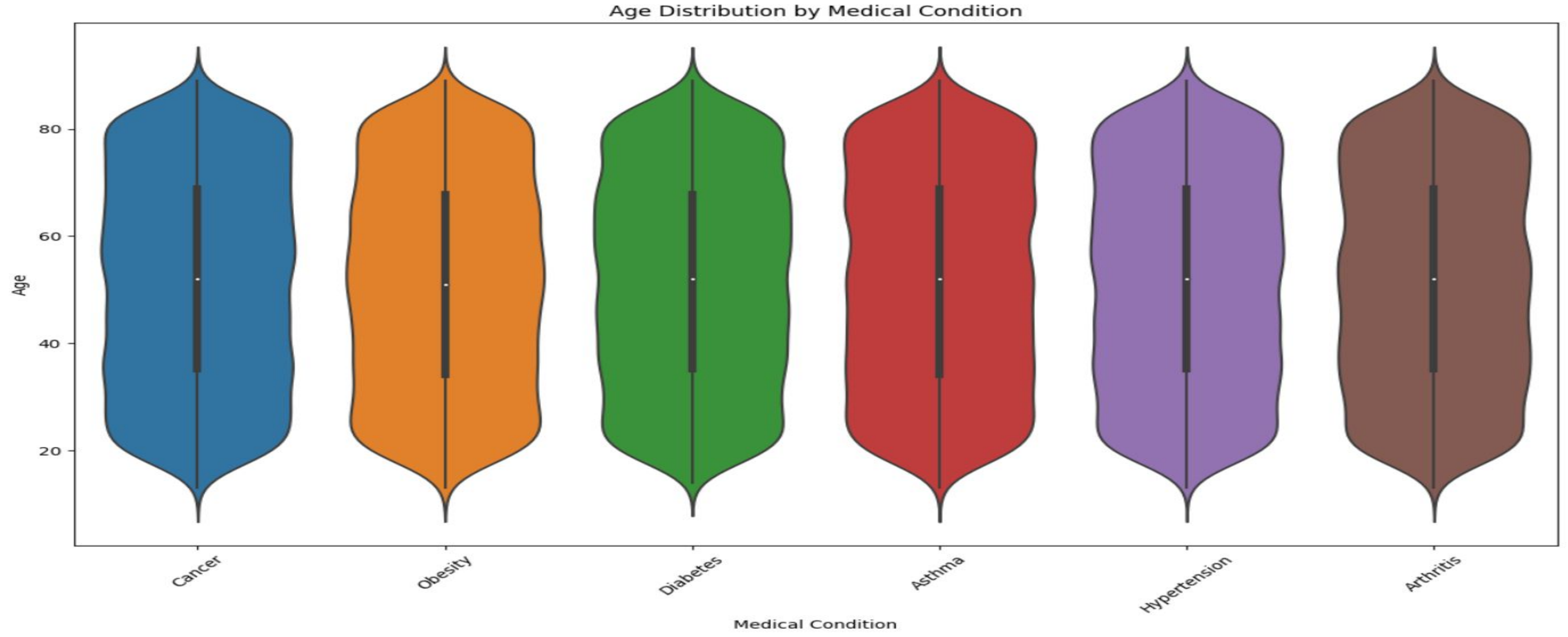
```python
heatmap_df = data_df[['Date of Admission','Medical Condition']].set_index('Date of Admission')
heatmap_df.head()
cleaned_df = heatmap_df.groupby('Date of Admission')['Medical Condition'].size().reset_index()
cleaned_df.set_index('Date of Admission', inplace=True)
cleaned_df.hvplot.heatmap(x='index.year', y='index.month', C='Medical Condition', cmap='blues')
```



```python
admissions_trend = data_df.resample('Y', on='Date of Admission').size()
admissions_trend.plot(title='Number of Admissions Over Time')
plt.xlabel('Date')
plt.ylabel('Number of Admissions')
plt.show()
```



```python
plt.figure(figsize=(12, 8))
sb.violinplot(x='Medical Condition', y='Age', data=data_df)
plt.title('Age Distribution by Medical Condition')
plt.xlabel('Medical Condition')
plt.ylabel('Age')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




THANK YOU