

```
In [14]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
# Set warning style
sns.set(style='whitegrid')
warnings.filterwarnings('ignore')
```

## Generate a Sample Healthcare Dataset

```
In [142]: dfpd.read_csv('healthcare_dataset.csv')

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096279	205	Emergency	2022-10-07	Aspirin	Normal
3	andrew watts	28	Female	A+	Diabetes	2020-11-18	Karin Wells	Hernandez Rogers and Vang.	Aetna	27909.702410	450	Urgent	2020-12-18	Ibuprofen	Abnormal
4	andrew NEI	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Medicare	14238.317614	458	Elective	2022-10-09	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	elizabeth JOHNSON	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2656.714952	417	Elective	2020-09-15	Penicillin	Abnormal
55496	Kyle PEREZ	61	Female	AB+	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	317	Elective	2020-02-01	Aspirin	Normal
55497	Katherine WANG	38	Female	B-	Hypertension	2020-07-13	Jon Jacobs DVM	and Mahoney Johnson Veterinary	UnitedHealthcare	27620.744717	346	Urgent	2020-08-10	Ibuprofen	Abnormal
55498	Jennifer JONES	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro	Medicare	32451.090358	321	Elective	2019-05-31	Ibuprofen	Abnormal
55499	JAMES GARCIA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.131172	448	Urgent	2024-04-29	Ibuprofen	Abnormal

55500 rows x 15 columns

```
In [140]: # Display first few rows of the dataset
print(df.head(5))

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096279	205	Emergency	2022-10-07	Aspirin	Normal
3	andrew watts	28	Female	A+	Diabetes	2020-11-18	Karin Wells	Hernandez Rogers and Vang.	Aetna	27909.702410	450	Urgent	2020-12-18	Ibuprofen	Abnormal
4	andrew NEI	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Medicare	14238.317614	458	Elective	2022-10-09	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	elizabeth JOHNSON	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2656.714952	417	Elective	2020-09-15	Penicillin	Abnormal
55496	Kyle PEREZ	61	Female	AB+	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	317	Elective	2020-02-01	Aspirin	Normal
55497	Katherine WANG	38	Female	B-	Hypertension	2020-07-13	Jon Jacobs DVM	and Mahoney Johnson Veterinary	UnitedHealthcare	27620.744717	346	Urgent	2020-08-10	Ibuprofen	Abnormal
55498	Jennifer JONES	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro	Medicare	32451.090358	321	Elective	2019-05-31	Ibuprofen	Abnormal
55499	JAMES GARCIA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.131172	448	Urgent	2024-04-29	Ibuprofen	Abnormal

55500 rows x 15 columns

```
In [140]: # Display first few rows of the dataset
print(df.head(5))

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096279	205	Emergency	2022-10-07	Aspirin	Normal
3	andrew watts	28	Female	A+	Diabetes	2020-11-18	Karin Wells	Hernandez Rogers and Vang.	Aetna	27909.702410	450	Urgent	2020-12-18	Ibuprofen	Abnormal
4	andrew NEI	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Medicare	14238.317614	458	Elective	2022-10-09	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	elizabeth JOHNSON	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2656.714952	417	Elective	2020-09-15	Penicillin	Abnormal
55496	Kyle PEREZ	61	Female	AB+	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	317	Elective	2020-02-01	Aspirin	Normal
55497	Katherine WANG	38	Female	B-	Hypertension	2020-07-13	Jon Jacobs DVM	and Mahoney Johnson Veterinary	UnitedHealthcare	27620.744717	346	Urgent	2020-08-10	Ibuprofen	Abnormal
55498	Jennifer JONES	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro	Medicare	32451.090358	321	Elective	2019-05-31	Ibuprofen	Abnormal
55499	JAMES GARCIA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.131172	448	Urgent	2024-04-29	Ibuprofen	Abnormal

55500 rows x 15 columns

```
In [140]: # Basic info about the dataset
df.info()

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096279	205	Emergency	2022-10-07	Aspirin	Normal
3	andrew watts	28	Female	A+	Diabetes	2020-11-18	Karin Wells	Hernandez Rogers and Vang.	Aetna	27909.702410	450	Urgent	2020-12-18	Ibuprofen	Abnormal
4	andrew NEI	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Medicare	14238.317614	458	Elective	2022-10-09	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	elizabeth JOHNSON	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2656.714952	417	Elective	2020-09-15	Penicillin	Abnormal
55496	Kyle PEREZ	61	Female	AB+	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	317	Elective	2020-02-01	Aspirin	Normal
55497	Katherine WANG	38	Female	B-	Hypertension	2020-07-13	Jon Jacobs DVM	and Mahoney Johnson Veterinary	UnitedHealthcare	27620.744717	346	Urgent	2020-08-10	Ibuprofen	Abnormal
55498	Jennifer JONES	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro	Medicare	32451.090358	321	Elective	2019-05-31	Ibuprofen	Abnormal
55499	JAMES GARCIA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.131172	448	Urgent	2024-04-29	Ibuprofen	Abnormal

55500 rows x 15 columns

```
In [140]: # Basic info about the dataset
df.info()

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096279	205	Emergency	2022-10-07	Aspirin	Normal
3	andrew watts	28	Female	A+	Diabetes	2020-11-18	Karin Wells	Hernandez Rogers and Vang.	Aetna	27909.702410	450	Urgent	2020-12-18	Ibuprofen	Abnormal
4	andrew NEI	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Medicare	14238.317614	458	Elective	2022-10-09	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	elizabeth JOHNSON	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2656.714952	417	Elective	2020-09-15	Penicillin	Abnormal
55496	Kyle PEREZ	61	Female	AB+	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	317	Elective	2020-02-01	Aspirin	Normal
55497	Katherine WANG	38	Female	B-	Hypertension	2020-07-13	Jon Jacobs DVM	and Mahoney Johnson Veterinary	UnitedHealthcare	27620.744717	346	Urgent	2020-08-10	Ibuprofen	Abnormal
55498	Jennifer JONES	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro	Medicare	32451.090358	321	Elective	2019-05-31	Ibuprofen	Abnormal
55499	JAMES GARCIA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.131172	448	Urgent	2024-04-29	Ibuprofen	Abnormal

55500 rows x 15 columns

```
In [140]: # Basic statistics for numerical columns
df.describe()

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
count	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000	55500.000000
mean	51.538459	25539.316097	301.134829	19.602454	14211.454431	115.243069	13.000000	-2008.492140	101.000000	25%	35.000000	13241.224652	202.000000	98%	52.000000
std	19.602454	14211.454431	115.243069	13.000000	-2008.492140	101.000000	25%	35.000000	13241.224652	202.000000	98%	52.000000	25539.316097	301.134829	115.243069
min	13.000000	-2008.492140	101.000000	25%	35.000000	13241.224652	202.000000	98%	52.000000	25539.316097	301.134829	115.243069	13.000000	-2008.492140	101.000000
max	89.000000	52764.276736	506.000000	75%	68.000000	37820.508436	401.000000	max	89.000000	52764.276736	506.000000	75%	68.000000	37820.508436	401.000000

```
In [140]: # Check for missing values
df.isnull().sum()

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096279	205	Emergency	2022-10-07	Aspirin	Normal
3	andrew watts	28	Female	A+	Diabetes	2020-11-18	Karin Wells	Hernandez Rogers and Vang.	Aetna	27909.702410	450	Urgent	2020-12-18	Ibuprofen	Abnormal
4	andrew NEI	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Medicare	14238.317614	458	Elective	2022-10-09	Penicillin	Abnormal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
55495	elizabeth JOHNSON	42	Female	O+	Asthma	2020-08-16	Joshua Jarvis	Jones-Thompson	Blue Cross	2656.714952	417	Elective	2020-09-15	Penicillin	Abnormal
55496	Kyle PEREZ	61	Female	AB+	Obesity	2020-01-23	Taylor Sullivan	Tucker-Moyer	Cigna	31457.797307	317	Elective	2020-02-01	Aspirin	Normal
55497	Katherine WANG	38	Female	B-	Hypertension	2020-07-13	Jon Jacobs DVM	and Mahoney Johnson Veterinary	UnitedHealthcare	27620.744717	346	Urgent	2020-08-10	Ibuprofen	Abnormal
55498	Jennifer JONES	43	Male	O-	Arthritis	2019-05-25	Kimberly Curry	Jackson Todd and Castro	Medicare	32451.090358	321	Elective	2019-05-31	Ibuprofen	Abnormal
55499	JAMES GARCIA	53	Female	O+	Arthritis	2024-04-02	Dennis Warren	Henry Sons and	Aetna	4010.131172	448	Urgent	2024-04-29	Ibuprofen	Abnormal

55500 rows x 15 columns

```
In [140]: # Distribution of medical conditions
df['Medical Condition'].value_counts()

In [183]: df

Out[183]:
```

	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
0	Bobby Jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18556.281306	328	Urgent	2024-02-02	Paracetamol	Normal
1	Leslie Terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	375	Emergency	2019-08-26	Ibuprofen	Inconclusive
2	Danny Smith	76	Female	A-	Obesity	202									



