- **Title:** Predictive Healthcare Insights: Patient Readmission Analysis
- **Subtitle:** End-to-End Data Analytics Workflow (Python → Azure → MySQL → Power BI)
- **Name: Tanmay Sharma**
- Role: Data/Business Analyst
- Date: 2025/08/27

Problem Statement

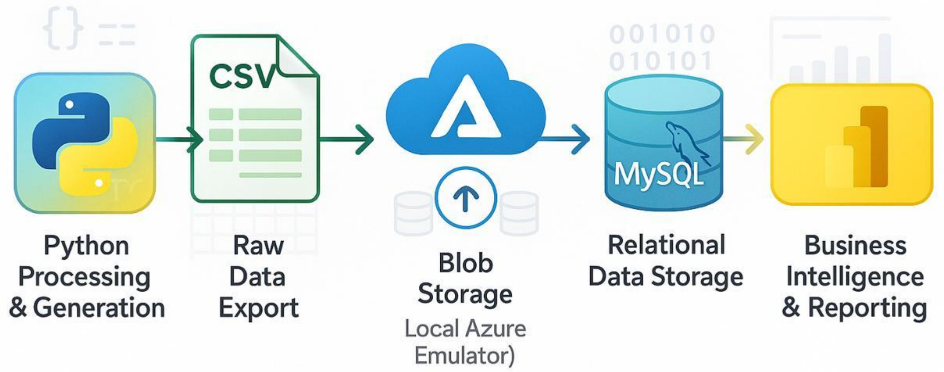
- **Content:**
  - Hospitals face **high readmission rates**, increasing cost and reducing patient care quality.
  - Goal: Predict **which patients are likely to be readmitted** and analyze patterns.

Objectives

- Predict patient readmissions using historical data
- Analyze **cost, length of stay**
- Build **interactive dashboards** for hospital management
- Store and retrieve data efficiently using **Azure + MySQL**

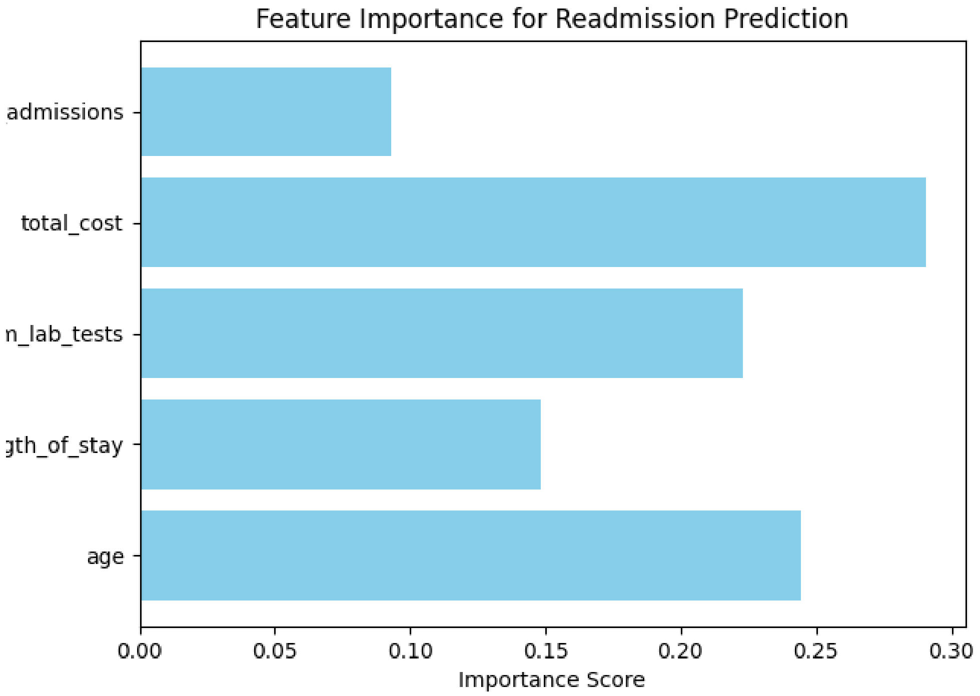
Data Sources

- **Hospital Admissions CSV:** Generated via Python (age, stay length, tests, cost, previous admissions, readmission)
- **Database:** MySQL (centralized storage)
- **Cloud Storage:** Azure Blob Storage (secure storage & integration)
- **Power BI Dashboard:** For visualization & decision-making



Python Data Processing

- **Steps:**
    1. Generate / Clean Dataset (hospital_admissions.csv)
    2. Feature Selection → age, length_of_stay, num_lab_tests, total_cost, prev_admissions
    3. Train-Test Split → test_size=0.2, random_state=42, stratify=y
    4. Pipeline → Imputer + Scaler + RandomForestClassifier
    5. Predictions → Actual vs Predicted + Readmission Probability
- **Insight:**
    - Feature importance shows **Length of Stay** and **Previous Admissions** most predictive
- **Visual:** Feature importance bar chart



```python
import pandas as pd
import numpy as np

# generate dummy healthcare admissions data
np.random.seed(42)
n = 1000

df = pd.DataFrame({
    "patient_id": np.arange(1, n+1),
    "age": np.random.randint(20, 90, size=n),
    "length_of_stay": np.random.randint(1, 15, size=n),
    "num_lab_tests": np.random.randint(5, 50, size=n),
    "total_cost": np.random.randint(2000, 50000, size=n),
    "prev_admissions": np.random.randint(0, 5, size=n),
    "target": np.random.choice([0,1], size=n, p=[0.7,0.3])  # 30% readmissions
})

df.to_csv("hospital_admissions.csv", index=False)
print("Dummy dataset saved as hospital_admissions.csv")
print(df.head())
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score, confusion_matrix, classification_report

# Load dataset
df = pd.read_csv("hospital_admissions.csv")

# Features & target
X = df[['age', 'length_of_stay', 'num_lab_tests', 'total_cost', 'prev_admissions']]
y = df['target']
```

Script.py

ML Evaluation

- **Metrics:**
  - ROC-AUC → Model discrimination power
  - Confusion Matrix → True Positive / False Positive insights
  - Classification Report → Precision, Recall, F1-Score
- **Insight:** 30% patients predicted likely readmission → helps hospital focus resources
- **Visual:** Confusion Matrix heatmap

```python
# Pipeline (impute + scale + model)
pipeline = Pipeline([
    ('imputer', SimpleImputer(strategy='median')),
    ('scaler', StandardScaler()),
    ('model', RandomForestClassifier(n_estimators=200, random_state=42, class_weight="balanced"))
])

pipeline.fit(X_train, y_train)

# Predictions
y_pred = pipeline.predict(X_test)
y_proba = pipeline.predict_proba(X_test)[:,1]

# Evaluation
print("\n📊 Model Performance:")
print("ROC AUC:", roc_auc_score(y_test, y_proba))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))

import matplotlib.pyplot as plt

model = pipeline.named_steps['model']
importances = model.feature_importances_
features = X.columns

plt.figure(figsize=(7,5))
plt.barh(features, importances, color="skyblue")
plt.xlabel("Importance Score")
plt.title("Feature Importance for Readmission Prediction")
plt.show()
```

SQL & MySQL Analysis

- **Queries:**
  - Average Length of Stay per Patient
  - Readmission Rate
  - Cost Analysis (Length of Stay vs Total Cost)
  - Gender-wise Readmission Rate
  - Monthly Admissions Trend
- **Insight:**
  - Readmission higher in patients with longer stay or frequent previous admissions

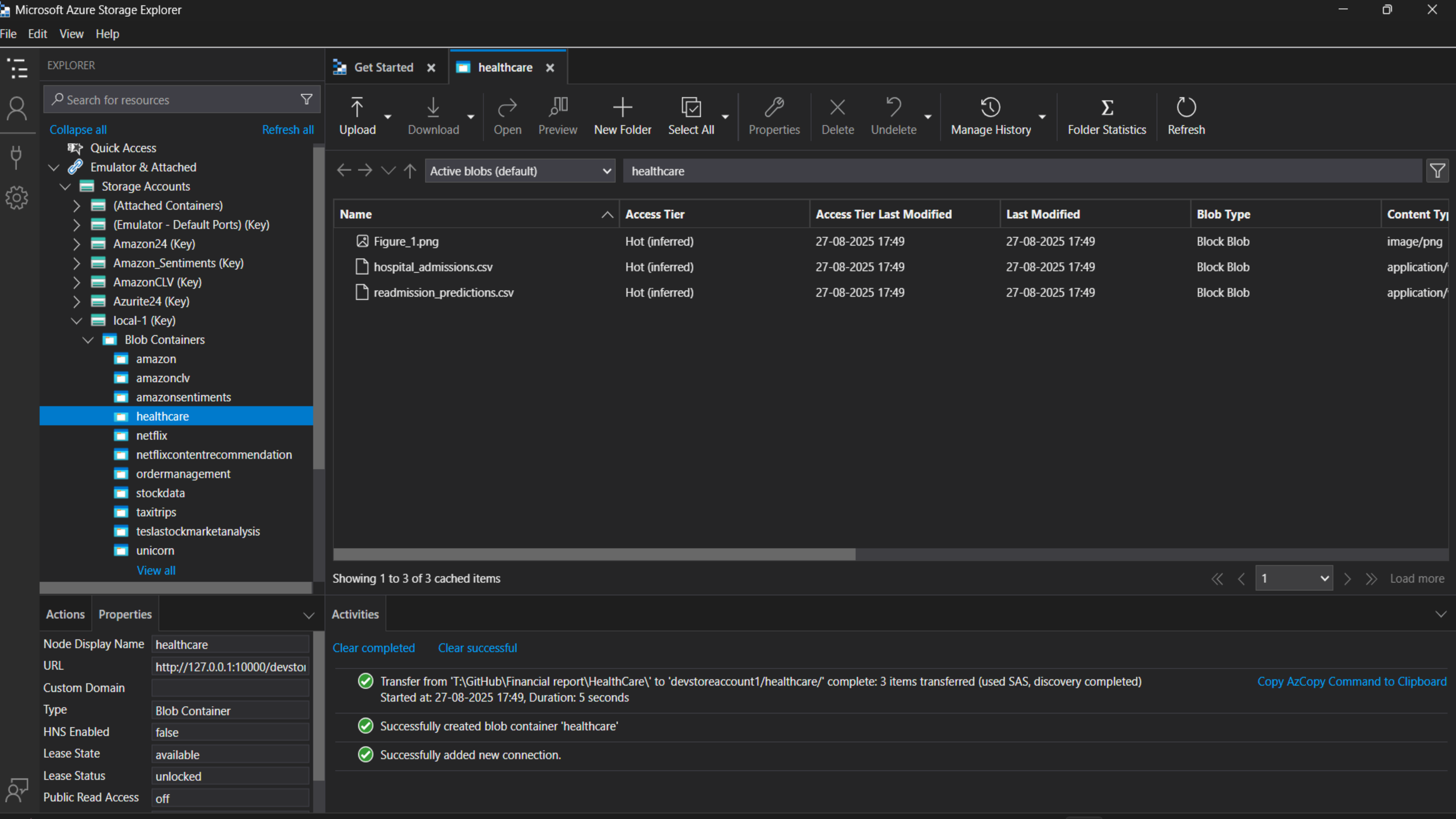
```sql
-- Monthly Admissions Trend
SELECT YEAR(admit_date) AS year,
       MONTH(admit_date) AS month,
       COUNT(*) AS total_admissions,
       ROUND(SUM(total_cost), 2) AS total_cost
FROM admissions
GROUP BY YEAR(admit_date), MONTH(admit_date)
ORDER BY year, month;
```

```sql
-- Gender-wise Readmission Rate
SELECT p.gender,
       COUNT(*) AS total_patients,
       SUM(a.readmitted) AS total_readmissions,
       ROUND(SUM(a.readmitted) * 100.0 / COUNT(*), 2) AS readmission_rate
FROM patients p
JOIN admissions a ON a.patient_id = p.patient_id
GROUP BY p.gender
ORDER BY total_readmissions DESC;
```

```sql
-- Patients with High Readmission Risk (Window Function)
WITH readmission_stats AS (
    SELECT a.patient_id,
           COUNT(*) AS total_admissions,
           SUM(a.readmitted) AS total_readmissions,
           ROUND(SUM(a.readmitted) * 100.0 / COUNT(*), 2) AS readmission_rate
    FROM admissions a
    GROUP BY a.patient_id
)
SELECT r.patient_id,
       p.name,
       r.total_admissions,
       r.total_readmissions,
       r.readmission_rate,
       RANK() OVER (ORDER BY r.readmission_rate DESC) AS risk_rank
FROM readmission_stats r
JOIN patients p ON r.patient_id = p.patient_id
ORDER BY r.risk_rank
LIMIT 10;
```

Azure Blob Storage

- **Content:**
  - Python CSV uploaded to Azure Blob Storage
  - Enables secure cloud access for dashboards
  - Integrated with Power BI for live updates
- **Insight:** Cloud storage ensures **scalable, central, and secure dataset**


Microsoft Azure Storage Explorer
File Edit View Help
EXPLORER
Search for resources
Collapse all
Refresh all
Quick Access
Emulator & Attached
Storage Accounts
(Attached Containers)
(Emulator - Default Ports) (Key)
Amazon24 (Key)
Amazon_Sentiments (Key)
AmazonCLV (Key)
Azurite24 (Key)
local-1 (Key)
Blob Containers
amazon
amazonclv
amazonsentiments
healthcare
netflix
netflixcontentrecommendation
ordermanagement
stockdata
taxitrips
teslastockmarketanalysis
unicorn
View all
Actions Properties
Node Display Name healthcare
URL http://127.0.0.1:10000/devstoı
Custom Domain
Type Blob Container
HNS Enabled false
Lease State available
Lease Status unlocked
Public Read Access off
Get Started
healthcare
Upload Download Open Preview New Folder Select All Properties Delete Undelete Manage History Folder Statistics Refresh
Active blobs (default)
healthcare
Name | Access Tier | Access Tier Last Modified | Last Modified | Blob Type | Content Typ
Figure_1.png | Hot (inferred) | 27-08-2025 17:49 | 27-08-2025 17:49 | Block Blob | image/png
hospital_admissions.csv | Hot (inferred) | 27-08-2025 17:49 | 27-08-2025 17:49 | Block Blob | application/
readmission_predictions.csv | Hot (inferred) | 27-08-2025 17:49 | 27-08-2025 17:49 | Block Blob | application/
Showing 1 to 3 of 3 cached items
1
Load more
Activities
Clear completed
Clear successful
Transfer from 'T:\GitHub\Financial report\HealthCare\' to 'devstoreaccount1/healthcare/' complete: 3 items transferred (used SAS, discovery completed)
Started at: 27-08-2025 17:49, Duration: 5 seconds
Copy AzCopy Command to Clipboard
Successfully created blob container 'healthcare'
Successfully added new connection.

Power BI Dashboard

- **Visuals Included:**
  - KPI Cards → Total Patients, Total Readmissions, Avg Length of Stay, Readmission Rate
  - Pie/Donut → Readmission vs No Readmission
  - Bar Chart → Readmission by Gender / Top 5 Costly Diagnoses
  - Line Chart → Monthly Admission Trends
  - Table → Patient-level Predictions (Actual vs Predicted + Probability)
- **Insight:**
  - Management can **identify high-risk patients and allocate resources proactively**

Total Patients

15

Total Readmissions

55.26

Avg_Length_of_Stay

5.20

AGE

38 | 72

Readmission Rate

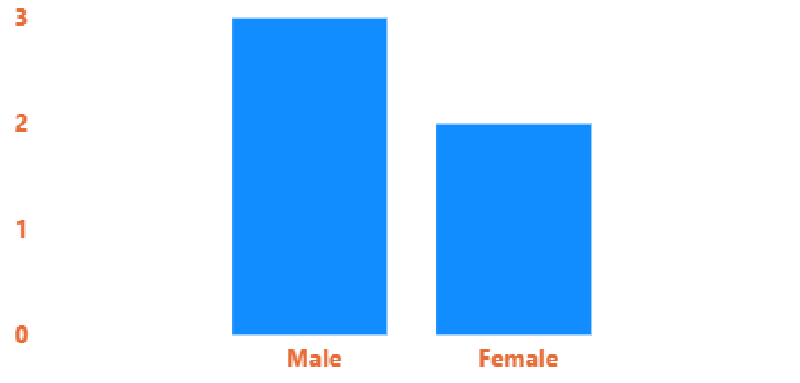
1.10K

"Predictive Healthcare Insights: Patient Readmission Analysis"

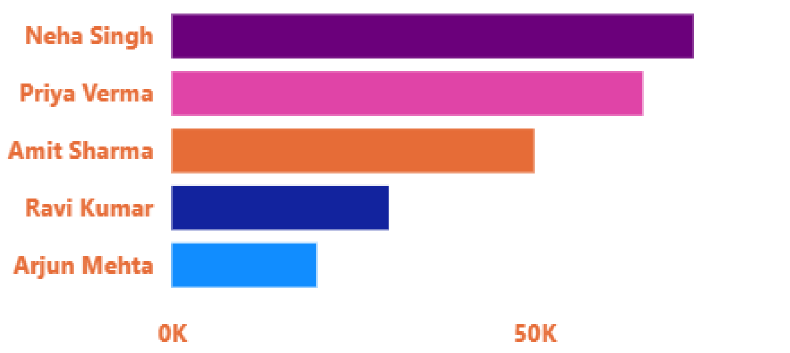
Admit_Date

- [ ] 01 May 2024
- [ ] 03 May 2024
- [ ] 05 May 2024

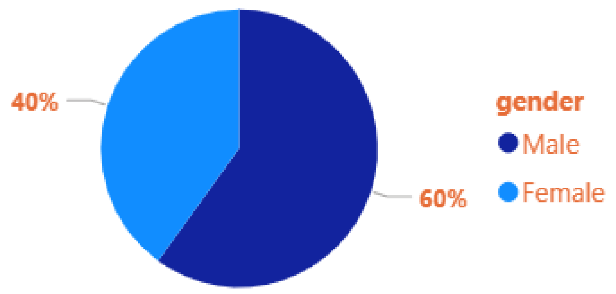
Readmission by Gender



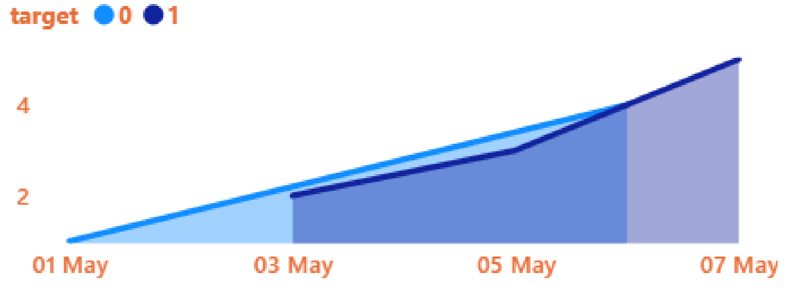
Top 5 expensive patients



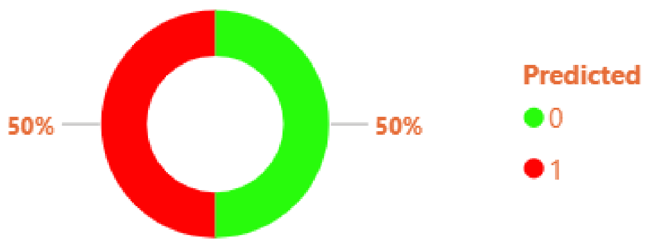
Gender Distribution



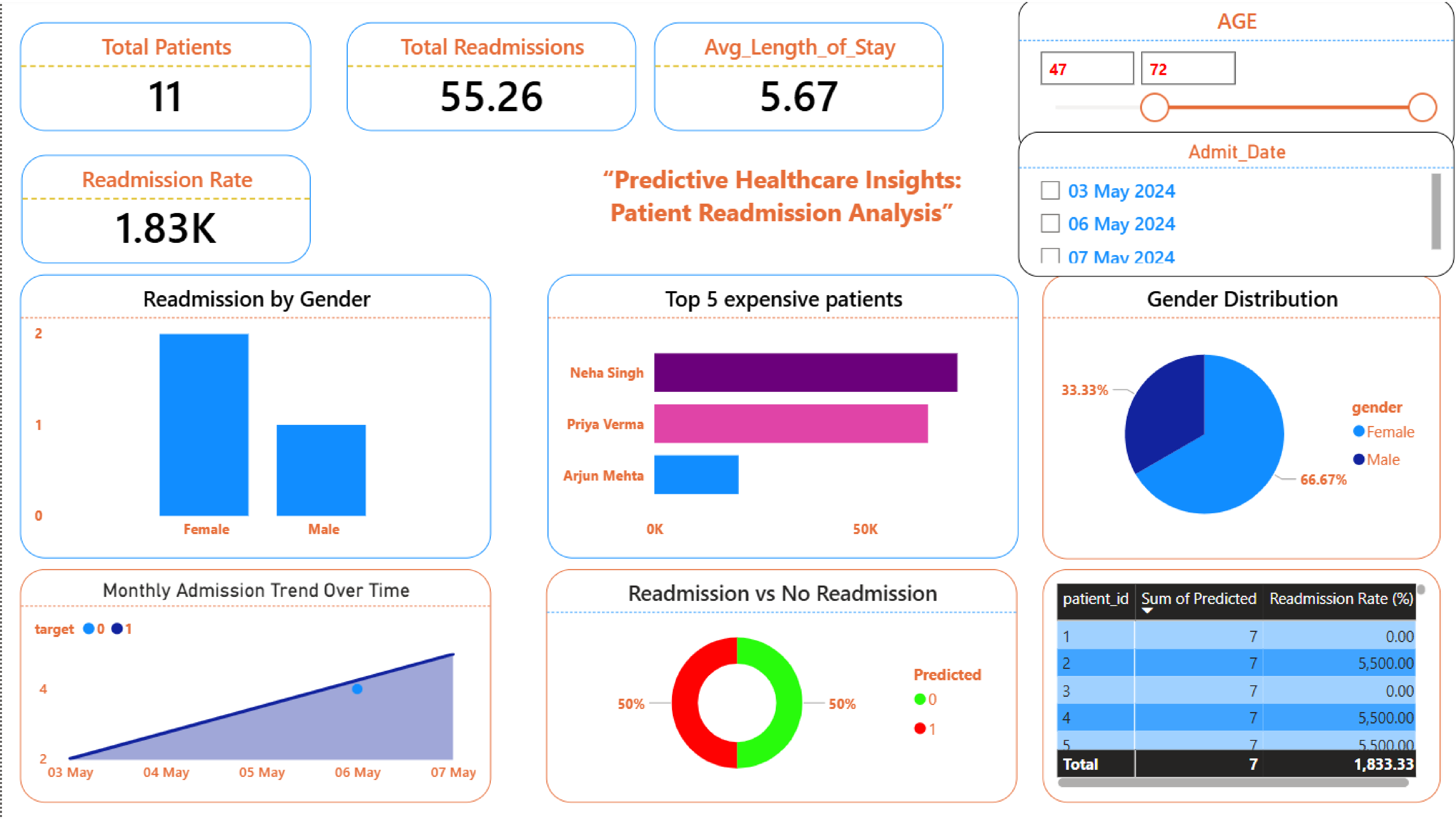
Monthly Admission Trend Over Time



Readmission vs No Readmission



patient_id	Sum of Predicted	Readmission Rate (%)
1	7	5,500.00
2	7	5,500.00
3	7	5,500.00
4	7	5,500.00
5	7	5,500.00
Total	**7**	**1,100.00**



Total Patients
11
Total Readmissions
55.26
Avg_Length_of_Stay
5.67
Readmission Rate
1.83K
AGE
47
72
Admit_Date
03 May 2024
06 May 2024
07 May 2024
"Predictive Healthcare Insights: Patient Readmission Analysis"
Readmission by Gender
2
1
0
Female
Male
Top 5 expensive patients
Neha Singh
Priya Verma
Arjun Mehta
0K
50K
Gender Distribution
33.33%
66.67%
gender
Female
Male
Monthly Admission Trend Over Time
target 0 1
4
2
03 May
04 May
05 May
06 May
07 May
Readmission vs No Readmission
50%
50%
Predicted
0
1
patient_id
Sum of Predicted
Readmission Rate (%)
1 7 0.00
2 7 5,500.00
3 7 0.00
4 7 5,500.00
5 7 5,500.00
Total 7 1,833.33

Total Patients	Total Readmissions	Avg_Length_of_Stay
7	55.26	4.50

AGE

47 63

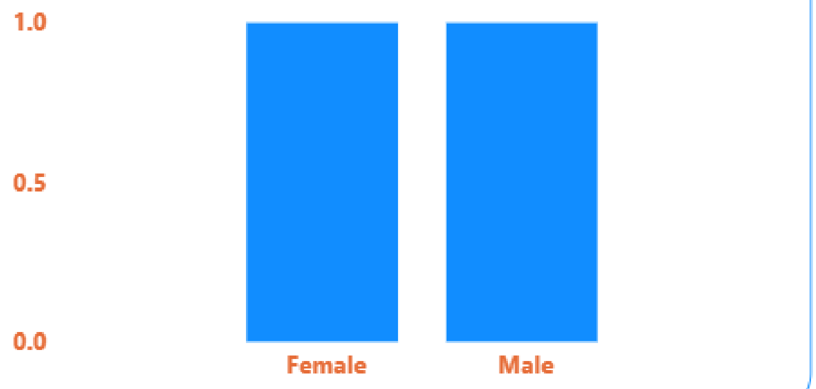
Readmission Rate
2.75K

“Predictive Healthcare Insights: Patient Readmission Analysis”

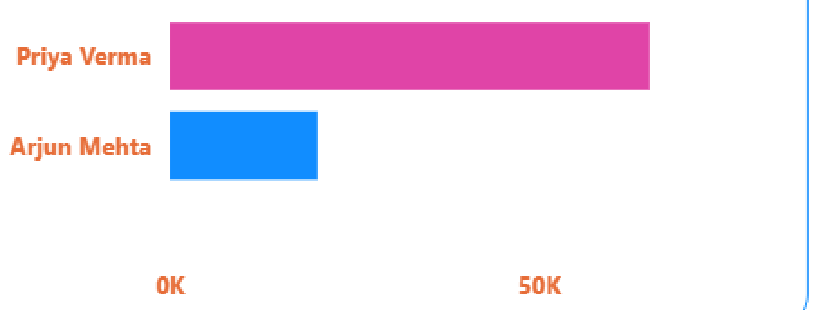
Admit_Date

- [ ] 03 May 2024
- [ ] 07 May 2024

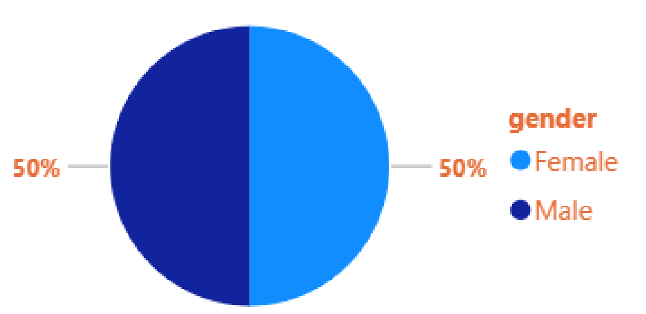
Readmission by Gender



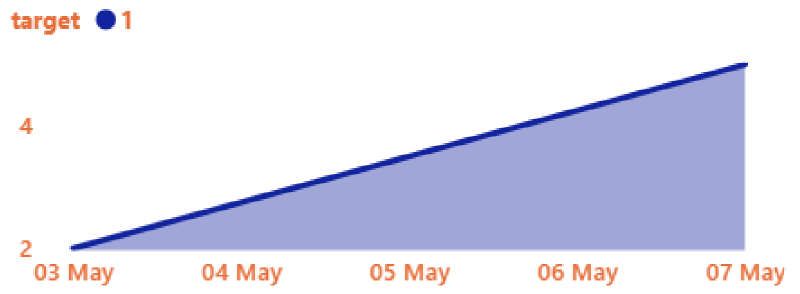
Top 5 expensive patients



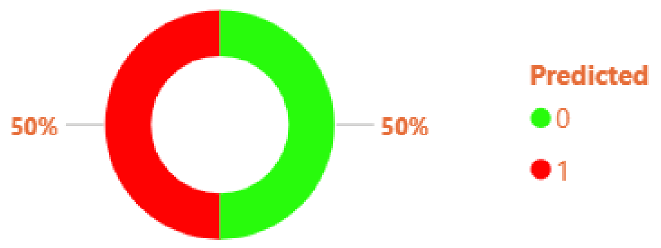
Gender Distribution



Monthly Admission Trend Over Time



Readmission vs No Readmission



patient_id	Sum of Predicted	Readmission Rate (%)
1	7	0.00
2	7	5,500.00
3	7	0.00
4	7	0.00
5	7	5,500.00
Total	**7**	**2,750.00**

Key Insights

- Readmission risk higher for patients with **longer stays and previous admissions**
- **Gender distribution** shows small differences in readmission patterns
- Cost correlates with length of stay → enables cost optimization
- Interactive dashboards help **real-time decision making**

Challenges & Learnings

- **Challenges:**
  - Imbalanced dataset → handled via stratify=y
  - Missing / inconsistent data → handled with SimpleImputer
  - Integrating Python → Azure → MySQL → Power BI pipeline
- **Learnings:**
  - End-to-end workflow mastery
  - Cloud integration (Azure Blob) for scalable analytics
  - MAANG-level data storytelling with dashboards

References / Tools

- Python: Pandas, Numpy, Scikit-learn, Matplotlib
- SQL: MySQL Workbench
- Cloud: Azure Blob Storage
- BI: Power BI
- Dataset: Synthetic Healthcare Data

Conclusion

- Built **predictive model** + **dashboard** for hospital management
- Enables **better patient care and resource optimization**
- Ready to extend for real-world healthcare deployment

Thank You/Let’s Connect:
LinkedIn: http://www.linkedin.com/in/
Git hub: https://github.com/Tanu272004
Website: https://tanmayportfolio52.wordpress.com/