**Data Analysis Project - Assignment 2**

**Student Name:** Youssef Tarek

**ID**: 231027091

**Task:** Analyzing Hospital Patient Data using Python

**1. Introduction**

In this assignment, I used Python to analyze a hospital dataset. The main goal was to understand patient demographics, identify common health conditions, and calculate how long patients usually stay in the hospital.

**2. Data Cleaning (Handling the Mistakes)**

Before I could get any results, I had to fix some problems in the data. This is an important step because "dirty data" gives wrong results.

- **Fixing Date Errors:** I found some rows where the Discharge Date was before the Admission Date. For example, patient **PN-7218608** was admitted in March 2025 but discharged 5 days earlier, which is impossible. I wrote code to filter out these rows.
- **Missing Information:** Some patients didn't have a diagnosis listed. Instead of deleting them, I filled the empty spots with **"Unknown"** so they wouldn't mess up the count.
- **Total Clean Records:** After cleaning, I was left with **4,850 valid records** to work with.

**3. Python Methodology**

I used the **Pandas** library to handle the data. The steps were:

1. Converted the date columns into a format Python can understand.
2. Calculated a new column called **"Stay"** (Discharge minus Admission).
3. Filtered the data to keep only valid stays (0 days or more).

**4. Key Findings & Results**

After running my Python script, I found the following:

- **Average Patient Age:** The average age of patients in this dataset is **47.3 years**.
- **Hospital Efficiency:** On average, patients stay in the hospital for about **5.5 days**.

- **Most Common Conditions:**
    1. Urinary Tract Infection (334 cases)
    2. Cholelithiasis (325 cases)
    3. Atrial Fibrillation (315 cases)
    4. Osteoarthritis (313 cases) *(Note: There were 338 cases marked as Unknown).*

```
Total valid records analyzed: 4850
Average Patient Age: 47.3 years
Average Hospital Stay: 5.5 days

Top 5 Diagnoses:
Diagnosis
Unknown                    338
Urinary Tract Infection    334
Cholelithiasis             325
Atrial Fibrillation        315
Osteoarthritis             313
Name: count, dtype: int64
```

**5. Conclusion**

Using Python made it easy to find and fix errors in the dates that I wouldn't have noticed manually. The analysis shows that the hospital handles a variety of cases, with an average stay of about 5 and a half days per patient.