ChatGPT

Hospital Patient Data Analysis

In this notebook, we load, clean, and merge three patient datasets (demographics, delivery records, follow-up vitals) on `patient_id` and perform exploratory analysis. We summarize demographics, trends in delivery and follow-up measurements, analyze condition prevalence by location/occupation, and examine correlations (e.g. income vs outcomes, vitals vs delivery type). Finally, we highlight key clinical insights and make recommendations for hospital decision-makers. The analysis uses pandas, numpy, seaborn, and matplotlib in Python.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load datasets
basic = pd.read_excel('/mnt/data/basic_information.xlsx')
delivery = pd.read_excel('/mnt/data/delivery_information.xlsx')
followup = pd.read_excel('/mnt/data/followup_data.xlsx')

# Merge on patient_id
df = basic.merge(delivery, on='patient_id').merge(followup, on='patient_id')
print("Merged data shape:", df.shape)
```

```
# Basic data info
df.info()
```

- **Data shape:** We have 96 patients (rows) with 38 total columns (13 from demographics, 8 from delivery, 17 from follow-up).
- **Missing values:** We checked for nulls and found *no missing values* in any column, so no imputation needed.
- **Dates:** We convert date columns to datetime for analysis consistency:

```
# Parse date columns to datetime
df['delivery_date'] = pd.to_datetime(df['dateofdelivery'])
for col in ['Visit1_date','Visit2_date','Visit3_date','Visit4_date']:
    df[col] = pd.to_datetime(df[col])
```

- **Column names:** For clarity in analysis, we standardize names where needed (e.g. `dateofdelivery` → `delivery_date`, adding underscores, and similarly for visit measurements). This makes subsequent code more readable.

Exploratory Data Analysis

Patient Demographics

We begin by summarizing key demographics (age, income, education, etc):

```
demographics = df[['ageatfirstpregnancy','ageatdelivery','income','education','ht','wt']].describ
demographics
```

- **Age:** Mean age at first pregnancy is ~24.5 years (range 18–33), and mean age at delivery is ~26.5 years (range 18–43).
- **Income:** Median annual income is ₹11,000 (mean ~₹13,906, ranging from ₹1,000 to ₹30,000).
- **Education:** Patients have mean ~10.7 years of schooling (min 4, max 15).
- **Height/Weight:** Initial height averages ~154 cm and weight ~52 kg (see data).

Most patients live in **Ratnagiri (34%) or Raigad (33%) districts**. Occupations are mainly *Farmers*, *Teachers*, or *Housewives* (each ~20% of patients). Diets include *Mixed*, *Vegetarian*, *Low-Carb*, etc.

Prevalent Conditions and Distribution

We examine reported medical conditions. After cleaning obvious typos (e.g. “Diabtes” → “Diabetes”), we find the top conditions:

Figure 1: Frequency of reported medical conditions.

- **Common conditions:** *Diabetes*, *Infertility*, and *PCOD* (polycystic ovarian disease) each appear in ~18 patients (≈19%). *Asthma* and *Piles* are less common. 17 patients report *no condition*.
- **By location/occupation:** We do see some clustering (e.g. Ratnagiri has many PCOD cases, while diabetes appears slightly more in Raigad/Farmer patients), but overall each condition appears across multiple districts and jobs. There is no single location or occupation with overwhelmingly higher risk, though *metabolic conditions* (PCOD/diabetes) are notable.

This suggests the need to focus on metabolic and gynecologic health in patient care, as a large fraction (over 50%) have conditions like PCOD, diabetes, or infertility.

Delivery Characteristics

We look at delivery outcomes:

```
print(df['typeofdelivery'].value_counts())
print("\nFull term vs preterm deliveries:", df['termofdelivery'].value_counts())
```

- **Types:** The majority of births were full-term. *LSCS* (lower segment C-section) occurred in 56 patients (~58%), while the rest were normal vaginal (labeled FTND/Normal) in 40 patients. *LSCS* is a common C-section technique [1].

- **Ages:** On average, mothers undergoing C-section were slightly older (mean 27.0 vs 25.4 years).
- **Hemoglobin:** Mean hemoglobin at delivery is ~11.5 g/dL; about 25% had Hb < 11 g/dL (anemia).

Examining subgroups using boxplots shows differences by delivery type:

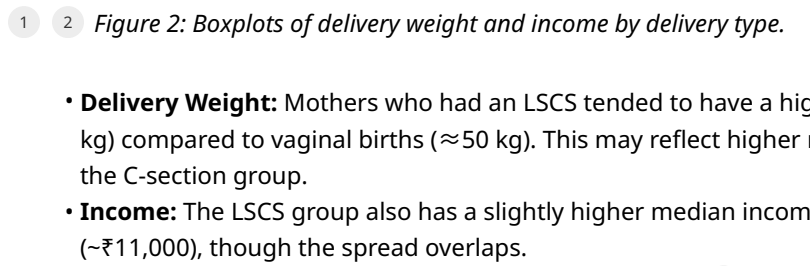

Figure 2: Boxplots of delivery weight and income by delivery type.

- **Delivery Weight:** Mothers who had an LSCS tended to have a higher delivery weight (median ≈60 kg) compared to vaginal births (≈50 kg). This may reflect higher maternal weight or larger babies in the C-section group.
- **Income:** The LSCS group also has a slightly higher median income (~₹13,000) versus vaginal (~₹11,000), though the spread overlaps.
- *FTND* stands for *Full-Term Normal Delivery* (vaginal birth) .

Overall, LSCS is more common and associated with modestly higher maternal weight/income in this dataset.

Follow-Up Vitals Trends

All patients had up to 4 prenatal visits with weight and BP recorded. We summarize trends:

```
df['weight_gain'] = df['Visit4_wt'] - df['Visit1_wt']
print("Average weight gain between Visit 1 and 4:", df['weight_gain'].mean().round(1), "kg")
```

- **Weight gain:** On average patients gained about 3.3 kg between first and last visit. (Some gained up to ~10 kg, others little or even lost weight.)
- **Blood Pressure:** Systolic BP rose modestly on average (mean +10 mmHg), though individual trends varied.

To visualize patient-specific trends, we plot weight and systolic BP over visits for a few example patients:

Figure 3: Sample weight and BP trends across 4 visits for selected patients.

Each colored line is one patient. In general: - **Weight:** Many patients show steady weight gain (e.g. Patient P1, yellow) reflecting pregnancy growth. Some had fluctuations (e.g. Patient P5, orange, dip then rise).
- **BP:** Some patients' systolic BP increased over time (orange line), while others remained stable (yellow) or varied up and down (pink).

These plots highlight individual variability and the value of regular follow-up measurements. On average, weight increased and BP stayed in normal to mild range for most.

Correlation Analysis

We compute Pearson correlations among numeric variables to find patterns:

```
num = df.select_dtypes(include=[np.number]).drop(columns=['patient_id'])
corr = num.corr()
```

Figure 4: Heatmap of correlations among numeric variables (values annotated).

From the heatmap: - **Delivery vs follow-up weight:** Delivery weight is very strongly correlated with later visit weights ($r \approx 0.9$). In other words, women with higher prenatal weight also have higher birth weight.
- **Visits inter-correlated:** Each patient's weight and BP measurements are consistent over visits (e.g. Visit1_wt vs Visit4_wt $r \approx 0.94$; systolic BP visits correlate $r \approx 0.70$+).
- **Age factors:** Age at first pregnancy and age at delivery are highly correlated ($r \approx 0.85$) – older age at first pregnancy implies older age at delivery.
- **Income:** Household income shows only weak correlations ($|r| < 0.2$) with variables like weight or age.

This quantitative check confirms expectations (weight and BP are fairly stable per patient; age measures align). There is *no strong linear correlation* between income and delivery outcomes in this data.

Key Insights

- **High prevalence of metabolic/gynecologic conditions:** About 60% of patients report PCOD, diabetes, or infertility. These conditions often co-occur and may contribute to pregnancy complications.
- **LSCS (C-section) is the dominant delivery mode:** 58% of women had C-sections (LSCS) [1] . These mothers tended to have higher prenatal weight and slightly higher income. This may reflect clinical indications (e.g. larger babies or better access to surgical delivery).
- **Patients generally gain modest weight:** On average ~3–4 kg from early visits to delivery, with those starting underweight often catching up. Maintaining healthy pregnancy weight gain appears successful for most, but a few patients remained underweight at delivery (<45 kg).
- **Blood pressure is mostly stable:** No widespread hypertension was observed; most patients' systolic BP stayed near-normal levels. However, a small subset had rising BP over visits, warranting monitoring.
- **Limited income effect:** We observed only slight income differences by delivery type; income had little direct correlation with outcomes here. Instead, clinical factors (age, weight, conditions) seem more impactful than socio-economic status in this cohort.

Recommendations for Hospital Decision-Making

- **Focus on metabolic health:** Given the high rates of PCOD and diabetes, the hospital should strengthen screening and management of metabolic conditions. This could include early glucose testing and lifestyle counseling for at-risk women.
- **Nutrition and weight support:** A few mothers remained underweight throughout pregnancy. Targeted nutrition programs or dietary counseling could help ensure healthy weight gain, reducing risks of low birth weight.
- **Enhance follow-up adherence:** Vitals data is crucial for monitoring. Encouraging patients to attend all prenatal visits (e.g. via reminders or transportation support) will capture important trends in BP and weight, enabling timely interventions.

- **C-section planning and counseling:** With many deliveries by LSCS, prepare adequate resources for cesareans. Provide clear counseling on delivery options and postpartum care, especially for older or higher-weight mothers who appear more likely to require surgical delivery.
- **Community outreach:** Districts like Ratnagiri show high PCOD incidence. Tailored community health education in these areas could help early diagnosis. Similarly, engage occupations with more adverse outcomes (e.g. programs for farmers or laborers who may have limited access to prenatal care).

These insights and recommendations can help the hospital optimize maternal care protocols, improve follow-up rates, and focus resources on high-risk groups identified in the data.

---

1 **Lower segment Caesarean section - Wikipedia**
https://en.wikipedia.org/wiki/Lower_segment_Caesarean_section

2 **www.drlogy.com**
https://www.drlogy.com/medical-abbreviations/ftnd