◆ Project Name: Doczy HealthNet

🎯 Objective:

To analyze and validate large healthcare datasets—particularly **DOFR (Division of Financial Responsibility)** and **Payment Conditions Output**—to ensure accuracy, automate data handling, and generate actionable insights through **Power BI dashboards** for stakeholders in the healthcare domain.

🛠 Tools & Technologies Used:

- **Excel (Advanced formulas):** VLOOKUP, HLOOKUP, conditional checks, financial calculations
- **Python:** Pandas, NumPy for automation and data cleaning
- **Power BI:** ETL (Power Query Editor), dashboard creation, data modeling using DAX
- **Textract:** For PDF data extraction into Excel
- **SQL:** For query optimization, validation, and error checking
- **Jupyter Notebook:** For exploratory data analysis (EDA)

🔁 End-to-End Process Flow

✅ 1. Data Extraction

- PDF data (DOFR & Payment Conditions Output) extracted by developers using **Textract**
- Converted and handed over in **Excel** format to the Data Analyst

---

✅ 2. Data Cleaning & Validation (in Excel)

- **DOFR Output (Row-wise Validation):**
  - Checked each row to verify accurate financial responsibility assignments.
- **Payment Conditions Output (Column-wise Validation):**
  - Ensured correctness of payment conditions, terms, and rates column-by-column.
- **Key Actions:**
  - Identified and removed **duplicate entries**
  - Located **missing values** using `VLOOKUP`, `HLOOKUP`
  - Applied formulas for **payment % calculations**, contract adjustments, etc.

✅ 3. Automation & Optimization

- Developed **scripts** (Python, SQL) to automate repetitive validation tasks
- Ensured fast and accurate checking of large data volumes

✅ 4. Data Import & Transformation in Power BI

- Imported cleaned Excel data into Power BI
- Performed additional transformations in **Power Query Editor**:
    - Removed extra spaces
    - Standardized formats
    - Merged DOFR and Payment tables
    - Created table relationships (Primary/Foreign keys)

✅ 5. Data Modeling with DAX

- Defined measures like:
    - Total payments
    - Claim approval rates
    - Contract compliance metrics

✅ 6. Dashboard Development in Power BI

- Created **interactive dashboards** with:
    - **KPIs:** Total claims, pending approvals
    - **Line Charts:** Financial trends over time
    - **Bar Charts:** Department-wise financial responsibility (DOFR)
    - **Tables:** Detailed payment breakdowns
- Added **filters/slicers** for departments, payment categories, contract type, etc.
- Used **conditional formatting** to highlight errors or missing data

---

✅ 7. Publishing & Sharing

- **Published** reports to Power BI Service
- Set up **auto-refresh** schedules

- Used **role-based access control** to ensure data privacy (HIPAA compliance)

✅ 8. Stakeholder Collaboration & Feedback

- Worked closely with:
  - Business teams
  - Financial analysts
  - Healthcare domain experts
- Refined dashboards based on feedback and changing regulations

---

💡 Challenges & Solutions in the Project

Challenge	Solution
Large, unstructured data	Cleaned using Python & SQL scripts, automated validation
Medical complexity (codes, contracts)	Collaborated with domain experts, built a data dictionary
Policy changes	Adjusted processing logic regularly, maintained up-to-date dashboards
Merging multiple data sources	Used Power Query and standardized formats for integration
Ensuring compliance (HIPAA)	Used anonymization and RBAC (Role-Based Access Control)

📌 Key Takeaways:

- Demonstrated expertise in **ETL, data validation, dashboarding**
- Delivered **95%+ accuracy** in financial validation with **no rework**
- Strong grasp of **US healthcare claims**, **contract management**, and **report automation**
- Direct impact on decision-making for **financial recovery**, **fraud detection**, and **policy compliance**