🩺 Health Insurance Cost Analysis

This project analyses a health insurance dataset to explore how different factors such as **age**, **BMI**, **smoking status**, and **number of children** affect medical insurance charges. The dataset was provided as part of the IBM Data Analyst course on Coursera.

📊 Project Overview

- **Dataset Source**: IBM Developer Skills Network – Coursera
- **Tools Used**: Python (Pandas, Matplotlib, Seaborn, scikit-learn), JupyterLite (Pyodide environment)
- **Goal**: Understand relationships between variables and build a simple regression model to predict insurance charges.

Features in the Dataset

Feature	Description
`Age`	Age of the individual (int)
`Gender`	Gender encoded (1 = Male, 2 = Female)
`BMI`	Body Mass Index
`No_Of_Children`	Number of children/dependents
`Smoker`	1 = Smoker, 0 = Non-smoker
`Region`	Coded region (1 to 4)
`Charges`	Annual insurance charges in dollars

📈 Key Visualizations

- Distribution of **Charges**
- Charges vs. **Age**
- Charges vs. **BMI**
- Boxplot: **Smoker** vs. Charges
- Heatmap: **Correlation matrix**

🔍 Key Insights

- **Smokers** have significantly higher insurance charges than non-smokers.
- **BMI** over 30 tends to increase costs, especially for smokers.
- **Age** has a linear relationship with charges.
- Number of children has a **weak influence** on charges.

🤖 Machine Learning

A **Linear Regression model** was used to predict insurance charges. Preprocessing included:

- Converting categorical variables to integers
- Scaling features using **`StandardScaler`**
- Optional polynomial feature expansion

📁 Files

- **`Health_Insurance.ipynb`:** Jupyter notebook with full code and analysis
- **`medical_insurance_dataset.csv`:** Raw dataset used in this project (not included on GitHub due to size, link in notebook)

How to Run

1. Clone or download this repo
2. Open `Health_Insurance.ipynb` in a Jupyter environment
3. Install requirements:

   ```
   pip install pandas matplotlib seaborn scikit-learn
   ```

4. Run all cells from top to bottom

Author

Vidya V.G.
IBM Data Analyst Certificate | Aspiring Data Analyst in the UK
LinkedIn Profile (optional)
GitHub Portfolio
