**Healthcare-Insurance-Analysis**

**Project Description:**

This project aims to predict patients' healthcare costs and identify the factors contributing to this prediction to help healthcare insurance providers make better decisions. The project involves collating data from different sources, cleaning and transforming the data, exploring the data, testing hypotheses, and building machine learning models. The project uses Excel and Python for data analysis and machine learning, and SQL for data management.

**Objectives:**

Predict patients' healthcare costs and identify the factors contributing to this prediction. Learn the interdependencies of different factors and comprehend the significance of various tools at various stages of the healthcare cost prediction process. Test hypotheses related to hospitalization costs, types of hospitals, types of cities, smoking, and heart issues. Develop and evaluate machine learning models using regression, random forest, and extreme gradient boosting for cost prediction.

**Technologies Used:**

Python - For data cleaning, transformation, and visualization/For data analysis and machine learning.

SQL - For data management and retrieval.

Tableau - For Data reporting.

**Conclusion:**

The project successfully predicted patients' healthcare costs and identified the factors contributing to this prediction. The project also tested hypotheses related to hospitalization costs, types of hospitals, types of cities, smoking, and heart issues. The project developed and evaluated machine learning models using regression, random forest, and extreme gradient boosting for cost prediction. The project used Python for data analysis and machine learning, Excel for data cleaning, transformation, and visualization, and SQL for data management and retrieval.