Insurance Costs Analysis Project

Data Analyst: Agustin A. Zavagnini

Purpose:

The purpose of this project is to explore and analyze a medical insurance dataset to understand how demographic and health-related variables (such as age, BMI, sex, smoking status, and number of dependents) influence individual insurance charges.

This project demonstrates core data-analysis skills, including data cleaning, exploratory data analysis, statistical reasoning, feature interpretation, and communication of insights for non-technical stakeholders.

Scope / Major Project Activities:

Activity	Description
1. Data Cleaning & Preparation	● Inspect the dataset for missing values, duplicates, inconsistent formats, and outliers. ● Standardize numerical and categorical fields to ensure analytical consistency. ● Engineer relevant features (e.g., BMI categories, smoker risk groups) to improve interpretability.
2. Exploratory Data Analysis (EDA)	● Analyze variable distributions using descriptive statistics and visualizations. ● Identify relationships between demographic/health factors and insurance charges. ● Detect patterns such as risk clusters, high-cost groups, and cost drivers.
3. Statistical Analysis & Modeling	● Apply correlation and regression techniques to quantify the impact of each variable on cost. ● Assess multicollinearity, significance levels, and effect sizes for explanatory power. ● Build a simple predictive model to estimate insurance charges based on customer profile.
4. Insights, Interpretation & Business Recommendations	● Transform statistical findings into actionable insights.

	• Identify which segments (e.g., smokers, high BMI individuals, older age groups) generate disproportionate cost increases. • Provide recommendations for risk-based pricing or health-improvement programs.

This project does not include:

- **Machine learning optimization:** hyperparameter tuning or advanced modeling is outside the project's scope.
- **Real patient data:** the analysis uses synthetic, publicly available insurance data and does not involve sensitive or confidential information.

Deliverables:

Deliverable	Description/Details
1. Analytical Report (Notebook)	A structured walkthrough including data preparation, visualizations, statistical analysis, interpretation of findings, and clearly articulated conclusions.