# Introduction

The insurance industry has long relied on data-driven decision-making, but in today's rapidly evolving digital landscape, it faces challenges in adapting to new technologies compared to other sectors, particularly banking. Unique obstacles such as complex underwriting rules, decentralized customer information, and regulatory compliance inertia hinder the industry's progress.

Despite these challenges, emerging technologies like AI and blockchain are revolutionizing insurance, with data analytics playing a central role. Four key factors driving the emergence of analytics in insurtech include the explosion of unstructured data, advances in AI and machine learning, real-time processing capabilities, and increased computing power.

This dataset provides valuable insights into risk underwriting in health insurance by examining various attributes of insured individuals, such as age, sex, BMI, number of children, smoking status, and region, alongside their insurance charges. With 1338 rows of data and no missing values, the dataset is ideal for exploratory data analysis, statistical analysis, hypothesis testing, and training linear regression models to predict insurance premium charges.

Proposed tasks include:

1. Exploratory Data Analysis

2. Statistical hypothesis testing

3. Statistical Modeling

4. Linear Regression