Hello！ 欢迎查看我的 Sample Project

这个项目旨在「通过数据分析与模型构建 来预测影响纽约市所有企业存活率的因素」

文件夹「Final Model Running」中记录了对于已有数据包括地区人口，企业营业状态，地区收入中位数，可达性（巴士，地铁，停车位）的清洗，合并，去重，降噪，等操作。
文件夹「Final Model Running」中的「Analysis Survival Rate for NYC businesses_Random Forest.ipynb」主要运用随机森林来预测可能影响纽约市企业存活率的因素。
文件夹「Final Model Running」中的「Analysis Survival Rate for NYC businesses_Naive Bayesian.ipynb」主要运用贝叶斯定理来预测同样的问题，建议首先查看上一个文件。

除了数据【文件夹「data」】和分析【文件夹「Final Model Running」】，本项目还有一个书面的学术报告「Final Report.pdf」和一个专业的学术海报「Academic Poster.pdf」可供在线查看。

感谢您的阅览！希望通过此参考项目能作为评估我的专业水平的参考提供价值！
另外 这个 repo 的其他文件也请尽情查阅！！

Hello! Welcome to my Sample Project.

This project aims to "predict the factors influencing the survival rate of all businesses in New York City through data analysis and model building."

The folder "Final Model Running" contains records of data cleaning, merging, deduplication, and noise reduction operations on existing data, including area population, business operational status, median area income, and accessibility (bus, subway, parking spaces).
In the folder "Final Model Running," the file "Analysis Survival Rate for NYC businesses_Random Forest.ipynb" mainly uses Random Forest to predict the factors that may affect the survival rate of businesses in New York City.
The file "Analysis Survival Rate for NYC businesses_Naive Bayesian.ipynb" in the same folder primarily uses the Naive Bayesian to predict the same issue. It is recommended that you first review the previous file.

In addition to the data [folder "data"] and analysis [folder "Final Model Running"], this project also includes a written academic report, "Final Report.pdf," and a professional academic poster, "Academic Poster.pdf," both available for online viewing.

Thank you for reviewing! I hope this reference project can be useful for evaluating my professional level.
Besides this, feel free to Check out other folders on this repo!!