

美国大选数据可视化项目报告

马舒童

10235501462

项目背景

在 2024 年美国大选的背景下，本项目旨在通过图表和地图将选民支持率、候选人信息、政党动态等数据直观地呈现给公众，帮助理解大选过程中的关键数据。

项目目标

本项目的目标是通过数据分析和可视化技术，帮助用户全面了解 2024 年美国大选的选举动态，具体目标和实现的功能如下：

- 展示选举结果：**通过交互式地图展示美国各州候选人支持率。支持点击不同州查看该州的具体支持率，候选人得票数等，支持交互功能（点击、悬停查看详细数据）。
- 候选人信息展示：**提供不同候选人的政见内容。
- 支持率动态更新：**展示候选人支持率随时间变化，支持用户选择具体时间点查看数据。
- 实时选举数据获取与显示：**实时更新选举数据，展示候选人在各州的得票数、支持率等信息。通过 API 接口从民调数据平台获取实时数据，并将数据传递给前端进行展示。使用 Dash 结合异步加载技术确保数据实时更新。
- 多维度选民行为分析：**分析不同选民群体（如按种族、地区分布等）对候选人支持率的影响，使用 Pandas 和 Scikit-learn 进行选民行为分析，进行数据建模和特征提取。通过聚类分析对选民群体进行划分，并展示不同群体的投票倾向。
- 筹款情况分析：**收集并展示候选人筹款数据，通过图表展示各候选人的筹款总额、筹款来源、候选人财政三维分析等。使用地图、饼图、三维图展示筹款情况，帮助分析资金对选举竞争的影响。

数据来源

本项目的数据来源广泛，包括官方公开的选举数据、民调数据、候选人提供的信息以及其他政治分析平台的统计数据。主要的数据来源如下：

- 政府公开数据
- 民调数据（RCP）
- 候选人和政党数据：通过各候选人官网、新闻发布会及相关政治新闻平台收集。

数据处理：

本项目对于收集到的数据进行了一系列预处理，具体操作如下：

- 数据整合：**收集的数据格式多样，需要统一标准化为统一格式。使用 Python 的 Pandas 库将所有数据整合到一个统一的结构中，方便后续处理和分析。
- 数据清洗：**去除缺失数据、重复数据和异常数据，确保数据的准确性。使用正则表达式和数据校验方法清洗文本数据和数值数据。
- 数据转换：**对原始数据进行转换，将其转化为适合可视化的格式。通过 API 获取实时数据时，将实时数据与历史数据合并，以便进行趋势分析。
- 数据分组与聚合：**从大规模数据中提取关键统计信息，提炼核心指标，将复杂的数据

结构转化为更直观、简洁的形式，构建适合可视化的结构，突出趋势与分布。

可视化设计与实现

本项目的可视化设计遵循清晰、易读的原则，主要包括以下几种展示方式：

1. 地图可视化

地图在本项目的可视化中多次使用，在主页面、州详情页面、实时投票页面、候选人筹款情况等页面均有使用。以主页面为例，使用 **Plotly** 绘制美国地图，通过不同的颜色来展示各州的支持率。地图呈现的是动态变化的支持率，用户可以通过交互选择不同的时间和州查看数据。每个州的详细信息可以通过鼠标悬停查看，包括候选人支持率、投票率等数据。

2. 词云图展示政策主张

通过词云图展示了两位候选人的政策主张，词云图中字体的大小和颜色直观地反映了关键词的重要性和出现频率，让观众能够快速了解候选人的政策重点。

3. 交互式展示

使用 **Dash** 和 **Plotly** 实现了交互式网页应用，用户可以根据日期、州、候选人等维度自定义查询，实时查看支持率、得票情况、选民参与率等信息。

该系统支持实时数据更新，确保数据的时效性。

4. 图表展示

1) 横向条形图

主页面利用横向条形图清晰直观地呈现选举结果，直观展示候选人的总得票数对比。

2) 条形图

主页面底部利用条形图对比政党选举结果。

3) 堆叠条形图

选民分析页面使用堆叠条形图展示了 **2024** 年美国大选中不同种族群体对两位候选人的支持比例。

4) 桑基图

选民分析页面使用桑基图展示不同受教育程度的群体支持两位候选人的比例。

5) 气泡图

选民分析页面用气泡图展示了不同区域类型（城市、郊区、农村）对两位候选人的支持比例。通过气泡的大小来表示支持的比例，可以直观地比较不同区域类型对两位候选人的支持强度。

6) 饼图

选民分析页面用两个饼图分别展示了男性和女性选民在不同年龄组别对两位候选人的支持比例。

7) 环形饼图

候选人筹款情况页面使用环形饼图，展示了候选人筹款来源的分布情况，直观地展示各筹款来源的比例，便于了解候选人筹款的主要来源。

8) 三维散点图

候选人筹款情况页面利用三维散点图展示候选人的筹款金额、支出金额和债务金额之间的关系。

数据洞察

一、选民结构与支持倾向

（一）种族因素显著影响支持分布

白人对特朗普的支持率较高，达到 0.6 左右，而黑人、拉丁裔、亚裔等少数族裔对哈里斯的支持更为明显，支持比例分别约为 0.8、0.6、0.4。这表明不同种族群体在政治立场和候选人选择上存在明显差异，种族因素在美国大选中是影响选民决策的重要变量。候选人若想赢得选举，需制定针对性策略，争取不同种族群体的支持，尤其要关注少数族裔群体不断增长的政治影响力。

（二）教育程度与支持关系复杂

拥有大学学位的选民对哈里斯的支持率相对较高，而未拥有大学学位的选民则更倾向于特朗普。这反映出教育背景在塑造选民政治倾向方面起着重要作用，不同教育程度群体的经济状况、社会价值观等差异可能导致其对候选人政策主张的不同偏好。

（三）地区差异导致支持分化

城市地区对哈里斯的支持较为集中，郊区选民对特朗普和哈里斯的支持相对均衡，农村地区则更倾向于特朗普。地域发展不平衡、经济结构差异以及文化氛围不同可能是造成这种支持分化的原因。

（四）性别年龄因素呈现多样特征

在不同性别年龄组别中，选民支持倾向存在差异。男性选民在 30 - 44 岁和 45 - 64 岁年龄段对特朗普的支持率较高，而女性选民在 18 - 29 岁和 30 - 44 岁年龄段对哈里斯的支持相对突出。年龄和性别因素与选民的生活经历、社会角色等密切相关，进而影响其政治选择。

二、筹款情况与选举影响力

（一）筹款金额反映竞选实力

候选人的筹款金额在一定程度上体现了其竞选实力和资源获取能力。筹款金额较高的候选人通常能够在竞选活动中投入更多资源，如广告宣传、组织竞选活动等，从而提高自身的知名度和影响力。

（二）筹款来源影响竞选策略

从筹款来源分布来看，个人捐赠在候选人筹款中占据重要比例，这意味着候选人需要积极争取广大民众的支持。同时，政党支持、其他委员会捐赠等也是重要的资金来源，反映了政治团体在选举中的影响力。

（三）筹款、支出与债务关系揭示竞选态势

筹款金额高的候选人往往支出也较高，但支出的合理性和效益更为关键。部分候选人可能因债务金额较高而面临资金管理压力，这可能影响其竞选活动的持续开展和策略调整。

三、选举结果动态与趋势

（一）实时投票展现选举动态变化

通过实时投票数据的可视化展示，我们可以清晰观察到选举日不同时间点各州投票结果的变化趋势。

（二）选举人票分布决定选举走向

选举人票在各州的分布情况对选举结果至关重要。特朗普在部分关键州的选举人票领先，最终赢得大选。这表明候选人在各州的竞选表现和选民支持率直接影响选举人票的归属，进而决定选举的胜负。

（三）政党在国会选举中的表现影响政治格局

在众议院和参议院选举中，共和党赢得国会控制权。这一结果将对美国未来的政治决策、政策制定和政府运作产生深远影响，政党在国会中的力量对比将在一定程度上决定政策议程的推进方向。

项目难点

1. 地图可视化的实现

难点：地图呈现大选数据时，如何确保不同州支持率的颜色能够准确反映选举动态。

解决方案：使用 Plotly 结合 GeoJSON 格式的美国州界数据，实现交互式地图，确保颜色渐变准确且清晰地展示支持率。

2. 实时数据更新与交互

难点：如何保证用户界面响应快速，并且实时更新数据。

解决方案：使用 Dash 和 React.js 结合，确保前后端的数据同步与交互，采用异步加载技术减少等待时间，提升用户体验。

3. 选民行为分析的复杂性

难点：如何分析多维度的数据（如年龄、性别、收入、地理位置等）对选民支持率的影响。

解决方案：通过聚类分析对选民数据进行建模，挖掘选民行为背后的模式，并进行多维度分析。

项目心得

在这个项目中，我不仅学到了数据科学的基础知识，还学到了如何将这些知识应用于实际的选举数据分析中。通过结合可视化工具、机器学习算法和前端开发技术，我体会到跨学科知识融合的重要性。

数据清洗和预处理是整个项目的基础环节，只有确保数据的准确性和一致性，后续的分析 and 可视化才能顺利进行。通过实践，我更加深刻地理解了数据预处理的重要性。

在前端展示部分，我认识到用户体验不仅仅取决于界面设计，还包括数据的加载速度、交互的流畅性等。在这一点上，动态更新和异步加载技术极大地提升了项目的用户体验。

通过项目，我不仅学到了许多新的技术工具（如 Dash、Plotly、Scikit-learn 等），还锻炼了在实际项目中解决问题的能力。