**数据可视化(CSE 578) 项目报告**

吴昊

[haowu35@asu.edu](mailto:Haowu35@asu.edu)

1. **项目的目标**

XYZ 公司使用数据来开发人员的营销档案，然后将这些档案出售给众多公司用于营销目的。你是 XYZ 的员工和数据分析团队的成员。你的团队刚刚获得一个新的项目，要与当地希望扩招学员的 UVW 大学合作。UVW 已经选择将薪资作为关键统计数据，以确定其学位课程的营销标准。你必须利用美国人口调查局提供的数据开发营销档案，并将 50,000 美元作为薪资的关键数据。对于薪资低于和超过 50,000 美元的人员，要评估许多关键变量，包括年龄、性别、教育背景、婚姻状态、职业等等。

例如，如果数据显示大多数薪资低于 50,000 美元的人员是 34 岁以下的单身男性，且具有高中文凭，则该大学可通过制定针对适合该人群的学费、课程内容、甚至是线上或线下课程，来向该人群销售课程。

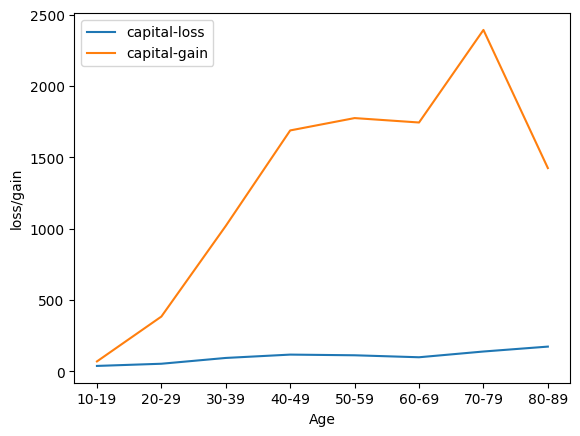
为了达到招生目标， UVW 的营销团队想要开发一个应用程序，以找到决定个人收入的因素。为此，一种方法是使用 XYZ 公司提供的美国人口调查局数据。对于在开发所提议的模型/应用程序过程中可以使用到的因素，营销团队希望对其进行分组。并希望利用该应用程序，根据不同输入参数值预测个人收入，以便调整适用于个人的营销方式。

1. **项目目标实现**

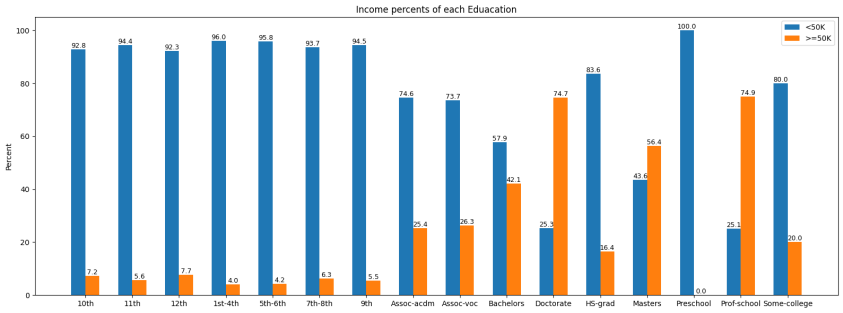
为了实现上述项目目标，采用了三个可视化实现，来帮助UVW大学实现对数据的分析和处理。

1. 不同年龄的人，分成10-19、 20-29、30-39、40-49、50-59、60-69、70-79、80-89八个年龄组对其在资本投资的损失(capital\_loss)和收益方面(capital\_gain)用折线图进行可视化显示。预期：在中间年龄段，有较大的资本收益。进行这个的分析主要是为了确定各个年龄段的投资损益情况，以便确定经济条件合适的年龄人群作为目标对象。
2. 对不同教育（education），了解收入（income）情况，用柱状图进行显示高低收入人所占百分比。以便更直观地展示受教育情况对收入的影响。
3. 使用词云显示不同原生国籍（native\_country）和收入（income）的关系。哪些原生国籍高收入占比高，哪些原生国籍低收入占比过高，以便学校据此做好招生。
4. **项目截图**

A.各年龄的投资损益图



B、各教育水平的高低收入百分比



1. 各原生国家人口在总体、高收入及低收入中的情况

（去除了United States）







1. **项目的有用性**
2. 不同年龄的人，分成10-19、 20-29、30-39、40-49、50-59、60-69、70-79、80-89八个年龄组对其在资本投资的损失(capital\_loss)和收益方面(capital\_gain)用折线图进行可视化显示。预期：在中间年龄段，有较大的资本收益。进行这个的分析主要是为了确定各个年龄段的投资损益情况，以便确定经济条件合适的年龄人群作为目标对象。

通过折线图我们看到，资本投资收益情况：60-69年龄组略有下降，在80-89年龄组大幅度下降。资本投资损失情况在50-59年龄组和60-69年龄组也有下降。

是否美国退休年龄和最早领取退休金年龄都是65岁。美国平均寿命77-78岁，考虑到进行资本投资的高收入人比例较多，平均寿命还会有一定的加大。

1. 对不同教育（education），了解收入（income）情况，用柱状图进行显示高低收入人所占百分比。以便更直观地展示受教育情况对收入的影响。

通过柱状图看到随着受教育程度的提高，50K年薪以上收入的比重都在增加（甚至PreSchool学前教育组没有50K以上占比为0）。

而且在接受过大学教育、学士、硕士、博士也是高收入占比逐步提高。

另外看到一个有意思的现象博士和教授的高收入者占比几乎是一样得（看来这是一个几乎平行或者几乎重叠的群体）。

因此教育是如此的重要（美国也是一个尊重知识的资本主义国家）。

1. 使用词云显示不同原生国籍（native\_country）和收入（income）的关系。哪些原生国籍高收入占比高，哪些原生国籍低收入占比过高，以便学校据此做好招生。

比较有意思的族群：

墨西哥在总的人口中占比最高，在收入<50K人口中占比也是最高，但是在50K以上人口中占比明显缩小（在排名前三只有），那么说明原生墨西哥的人口低收入者较多，高收入者较少。

另外在50K以上的高收入人口中菲律宾、德国、意大利、英格兰、中国等亚洲和欧洲国家人口占比比在总体人口占比中要高。时UVW大学招生特别要关注的人群。

萨尔瓦多、波多黎各等中南美洲的人口明显在低收入占比高于在总提人群中的占比。