Background Introduction

For many companies, especially Internet companies, they will attract others to log in to their website homepage and register members by advertising or occupying search engines. The ratio of actual signups to pageviews is called the conversion rate. And my project is to build a model and predict the conversion rate from an existing data set.

This program can provide companies with the means to analyze customer conversion rates. Companies can infer user preferences by providing back-office data from websites or apps, so as to find ways to improve user conversion rates. During the project, the dataset is normalized so that additional analysis parameters can be added more easily. In this way, the project can be applied to more situations.

1 Motivation

The main goal of this project is to build an appropriate model to assist the company in analyzing its own customer conversion rate. The main task is to search relevant data from the Internet and social media, standardize the data, build a model and complete the analysis.

2 Dataset Introduction

The dataset comes from Github, it is from 2017 about a site. The features have country, age, source, new_user, source, total_pages_visited and converted.

3 Basic data content

```
data = pd.read_csv('/content/drive/MyDrive/CS506_proposal/conversion.csv')
df = pd.DataFrame(data)
df.head()
```

	country	age	new_user	source	total_pages_visited	converted
0	UK	25	1	Ads	1	0
1	US	23	1	Seo	5	0
2	US	28	1	Seo	4	0
3	China	39	1	Seo	5	0
4	US	30	1	Seo	6	0

```
print("there are " + str(df.isnull().T.any().sum()) + " rows with at least one empty value")
```

```
there are 0 rows with at least one empty value
```

Figure 1: Average Age

4 Data Analysis

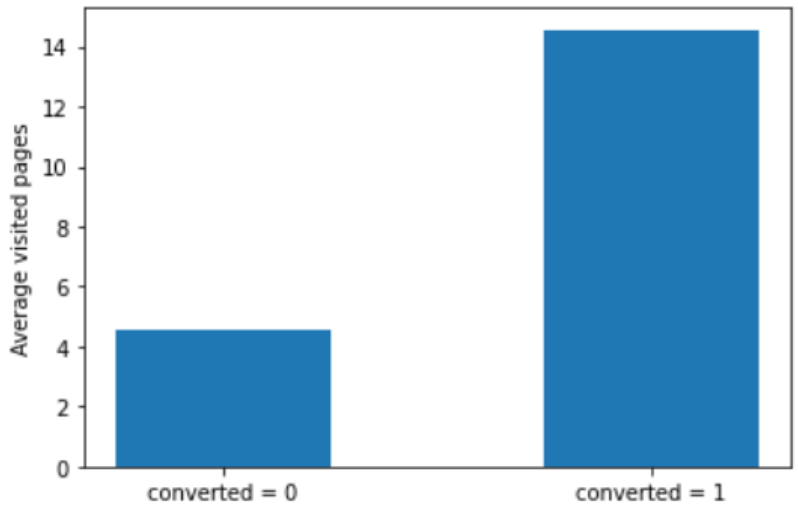

Figure 2: Average pages

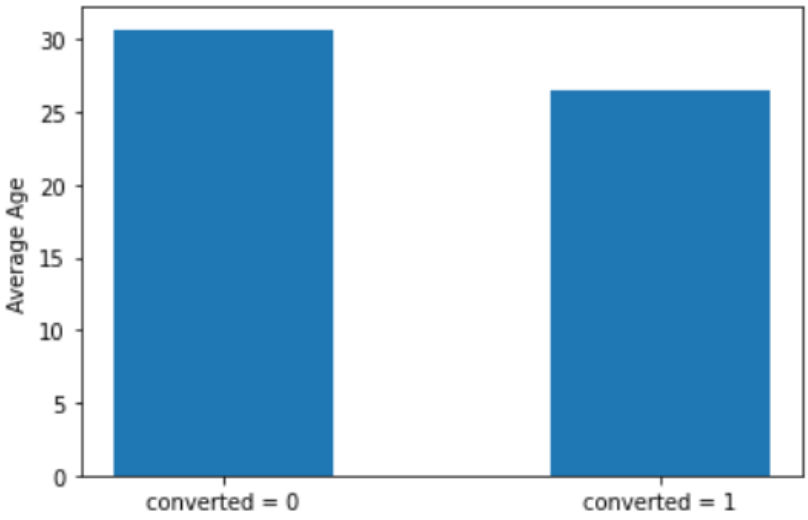

Figure 3: Average Age

5 Question to be Answered

The first question is what is the association between conversion rate and the user profile? In the process of my data processing, the pages visited have a great influence on the conversion

results, and most of the people who are converted have more pages. Age also plays a role, with converted people being on average younger than non-converted ones. The impact of new_user has not been counted, and country and source also have an impact on the results.