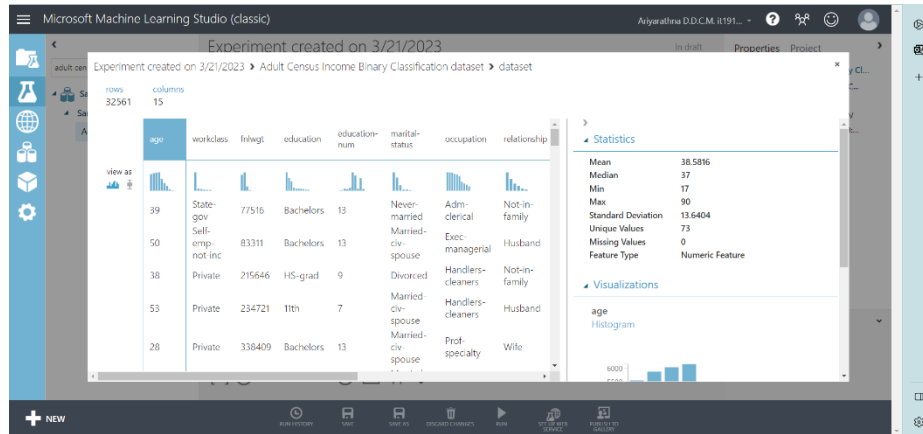


# 1. Adult Census Income Binary Classification Dataset

2.

a)



b)

1)

- Age
- fnlwgt
- education-num
- capital-gain
- capital-loss
- hours-per-week

2)

- Workclass – drop column
- Occupation – drop column
- Native-country – drop row

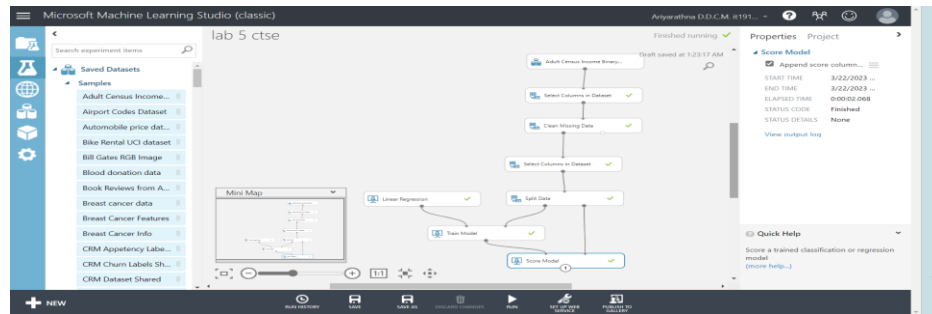
3)

- Workclass
- Education
- Marital-status
- Occupation
- Relationship
- Race
- Sex

4)

- Age

- ii. Fnlwgt
- iii. education-num
- iv. capital-gain
- v. capital-loss
- vi. hours-per-week



The screenshot shows the 'Scored dataset' view in Microsoft Machine Learning Studio (classic). The dataset has 7994 rows and 14 columns. The columns are: relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country, income, and Scored Labels. The 'Scored Labels' column contains values ranging from -0.018983 to 0.266677. The 'Statistics' and 'Visualizations' panes are also visible on the right.

relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income	Scored Labels
Own-child	White	Male	0	0	20	United-States	<= 50K	-0.018983
Unmarried	Black	Female	0	3004	35	United-States	> 50K	0.678403
Not-in-family	White	Male	0	0	45	Portugal	<= 50K	-0.068794
Own-child	Black	Male	0	0	35	United-States	<= 50K	0.014896
Other-relative	White	Female	0	0	25	United-States	<= 50K	-0.013464
Husband	Amer-Indian-Eskimo	Male	0	0	40	United-States	<= 50K	0.266677