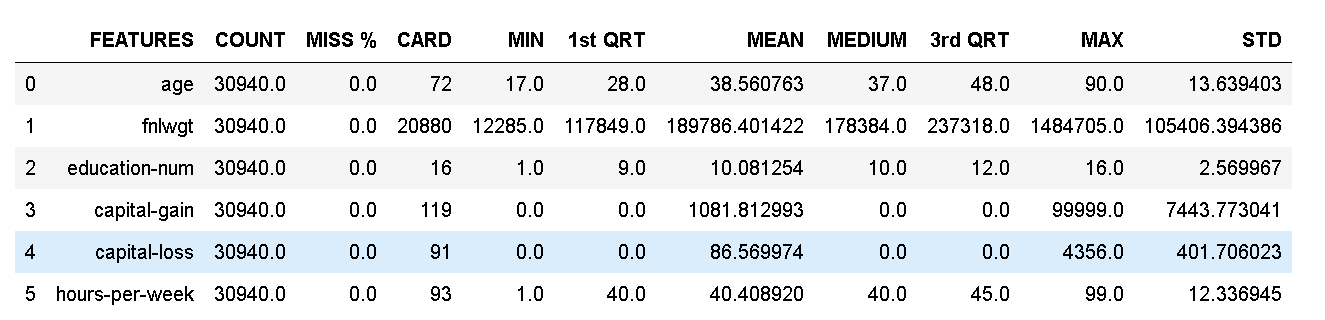
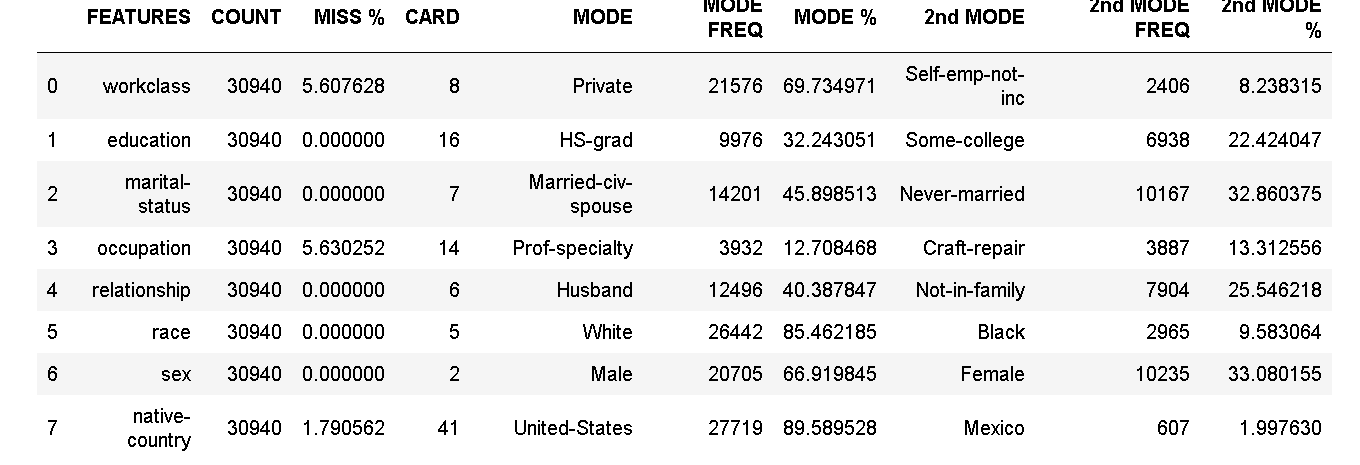
**Analysis report**





Missing value analysis:

Features workclass 5.6% missing values.

Possible solution: leave it as it is, document it in the data quality plan, will use imputation

Feature occupation 5.63% missing values

Possible solution: leave it as it is, document it in the data quality plan, very similar to workclass, may relate to workclass which is not complete. Will use imputation.

Features native-country have1.79% missing values.

Possible solution: leave it as it is, document it in the data quality plan. Will use imputation

Cardinality analysis:

Workclass, marital-status, relationship, race, sex all have cardinality <10

Decision: no issues

Outliers analysis

Capital-gain and capital-loss 1st, 2nd, 3rd QRT all is 0, that means 0 is most values in capital-gain and capital-loss.

Decision: we will only use mean, max and std from capital-gain and capital-lose.

Age, fnlwgt, capital-gain, capital-lose maximum values is unusually high values or very low values. We use clamp transformation that clamps all values above an upper threshold and below a lower threshold to these threshold values.

Age: lower = 38.56 – 13.64\* 2 =11.28 use 11 upper: 38.56+13.64\*2 = 65.84 use 66

Fnlwgt: lower = 189786 – 105406\*2 = -21026 use 0 upper= 189786 + 105406\*2 = 400598

Capital-gain: lower = 1082 – 7444\*2 = -13806 use 0 upeer=1082+7444\*2 = 15970

Capital-loss: lower = 87-402\* 2 use 0 upper = 87 + 402\*2 = 895

Hours-per-week: lower = 40-12\*2 = 16 upper = 40+12\*2 = 64

**Data Quality Plan**

|  |  |  |
| --- | --- | --- |
| Feature | Data quality Issues | Potential Handing Strategies |
| Workclass | Missing values(5.6%) | Imputation(Mode private) |
| Occupation | Missing values(5.63%) | Imputation(mode prof-specialty) |
| Native-country | Missing value(1.79%) | Imputation(mode united-states) |
| Capital-gain | Outilers(low or high) | Clamp transformation  (manual: 0,15970) |
| Capital-loss | Outilers(low or high) | Clamp transformation  (manual: 0,895) |
| Hours-per-week | Outilers(low or high) | Clamp transformation  (manual: 16,64) |
| age | Outilers(high) | Clamp transformation  (manual: 11,66) |
| fnlwgt | Outilers(low or high) | Clamp transformation  (manual: 0,400598) |