**Determining algorithmic bias in case of gender**

1. **Analysis of overall Data**
2. **With SSL\_SCORE and SEX\_CODE\_CD**
3. Without PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values on overall data for Females and Males without PSM** | | | | |
| **Average SSL\_SCORE value of Females** | **Average SSL\_SCORE value of Males** | **Difference between Average SSL\_SCORE values of Females and Males** | **T-Score** | **P-Value** |
| 283.4904 | 278.7586 | 4.73 | 23.514 | < 2.2e-16 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values on overall data for Females and Males with PSM** | | | | |
| **Average SSL\_SCORE value of Females** | **Average SSL\_SCORE value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | **P-Value** |
| 283.4904 | 279.5593 | 3.9311 | 13.398 | < 2.2e-16 |

1. With PSM (SSL\_SCORE and SEX\_CODE\_CD)
2. **With SSL\_RANK and SEX\_CODE\_CD**
3. Without PSM (SSL\_SCORE\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values on overall data for Females and Males without PSM** | | | | |
| **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| 190375.6 | 200479.6 | 10104 | -24.366 | < 2.2e-16 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values on overall data for Females and Males with PSM** | | | | |
| **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| 190375.6 | 194233.4 | 3857.8 | -6.9309 | 4.197e-12 |

1. With PSM (SSL\_RANK and SEX\_CODE\_CD)
2. **Analysis of records with age brackets for entire data**
3. **With SSL\_SCORE and SEX\_CODE\_CD**
4. Without PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males without PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE** **value of Females** | **Average SSL\_SCORE** **value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 342.2108 | 350.2799 | 8.06 | -49.131 | < 2.2e-16 |
| 20-30 | 309.4214 | 313.3934 | 3.97 | -36.06 | < 2.2e-16 |
| 30-40 | 266.7126 | 268.3676 | 1.65 | -13.328 | < 2.2e-16 |
| 40-50 | 222.6254 | 223.6764 | 1.051 | -6.7228 | 1.823e-11 |
| 50-60 | 181.2069 | 181.2420 | 0.0351 | -0.16406 | 0.8697 |
| 60-70 | 137.9924 | 138.8886 | 0.8962 | -1.8754 | 0.06091 |
| 70-80 | 91.94444 | 91.59264 | 0.3518 | 0.26642 | 0.7902 |

1. With PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males with PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE** value **of Females** | **Average SSL\_SCORE** **value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 342.2108 | 363.1329 | 20.9221 | -79.143 | < 2.2e-16 |
| 20-30 | 309.4214 | 325.9212 | 16.4998 | -97.496 | < 2.2e-16 |
| 30-40 | 266.7126 | 276.3343 | 9.6217 | -54.757 | < 2.2e-16 |
| 40-50 | 222.6254 | 231.4640 | 8.8386 | -42.672 | < 2.2e-16 |
| 50-60 | 181.2069 | 188.5393 | 7.3324 | -21.429 | < 2.2e-16 |
| 60-70 | 137.9924 | 147.9133 | 9.9209 | -13.485 | < 2.2e-16 |
| 70-80 | 91.94444 | 105.37302 | 13.4285 | -7.6504 | 4.351e-13 |

1. **With SSL\_RANK and SEX\_CODE\_CD**
2. Without PSM (SSL\_SCORE\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males without PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 48542.64 | 41384.53 | 7158.11 | 32.978 | < 2.2e-16 |
| 20-30 | 139338.8 | 130236.6 | 9102.2 | 30.882 | < 2.2e-16 |
| 30-40 | 250595.8 | 246601.1 | 3994.7 | 14.561 | < 2.2e-16 |
| 40-50 | 324864.6 | 323258.7 | 1605.9 | 7.033 | 2.077e-12 |
| 50-60 | 370539.5 | 370164.0 | 375.5 | 1.9248 | 0.05428 |
| 60-70 | 392940.2 | 392392.7 | 547.5 | 3.9083 | 9.541e-05 |
| 70-80 | 397982.8 | 397959.0 | 23.8 | 0.48201 | 0.6303 |

1. With PSM (SSL\_SCORE\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males with PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 48542.64 | 29092.52 | 19450.12 | 77.669 | < 2.2e-16 |
| 20-30 | 139338.83 | 99364.13 | 39974.7 | 101.39 | < 2.2e-16 |
| 30-40 | 250595.8 | 227977.9 | 22617.9 | 55.063 | < 2.2e-16 |
| 40-50 | 324864.6 | 311951.5 | 12913.1 | 39.815 | < 2.2e-16 |
| 50-60 | 370539.5 | 362156.8 | 8382.7 | 23.721 | < 2.2e-16 |
| 60-70 | 392940.2 | 388701.4 | 4238.8 | 11.257 | < 2.2e-16 |
| 70-80 | 397982.8 | 397325.1 | 657.7 | 5.5866 | 9.469e-08 |

1. **Analysis of first 1000 records without age brackets**
2. **SSL\_SCORE and SEX\_CODE\_CD**
3. Without PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values for first 1000 data for Females and Males without PSM** | | | | |
| **Average SSL\_SCORE value of Females** | **Average SSL\_SCORE** **value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | **P-Value** |
| 464.5952 | 466.3539 | 1.7587 | -0.52602 | 0.6015 |

1. With PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values on overall data for Females and Males with PSM** | | | | |
| **Average SSL\_SCORE** **value of Females** | **Average SSL\_SCORE** **value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | **P-Value** |
| 464.5952 | 475.7857 | 11.1905 | -2.486 | 0.01496 |

**B) SSL\_RANK and SEX\_CODE\_CD**

1. Without PSM (SSL\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values for first 1000 data for Females and Males without PSM** | | | | |
| **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| 513.2857 | 479.9729 | 33.3128 | 0.68575 | 0.4964 |

1. With PSM (SSL\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **T-test values on overall data for Females and Males with PSM** | | | | |
| **Average SSL\_SCORE\_RANK value of Females (A)** | **Average SSL\_SCORE\_RANK value of Males (B)** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| 513.2857 | 334.7857 | 178.5 | 2.7926 | 0.006521 |

1. **Analysis of first 1000 records with age brackets**
2. **SSL\_SCORE and SEX\_CODE\_CD**
3. Without PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males without PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE** **value of Females** | **Average SSL\_SCORE** **value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 462.3939 | 467.1624 | 4.7685 | -1.324 | 0.1941 |
| 20-30 | 472.6667 | 463.0511 | 9.615 | 1.1791 | 0.27 |

\*30-40: Only 6 records are present in this age group and their gender is Male and no female records are present

\*40-80: No records of this age group present in first 1000.

1. With PSM (SSL\_SCORE and SEX\_CODE\_CD)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males with PSM** | | | | | | | |
| **Age Brackets** | **Average SSL\_SCORE** **value of Females** | **Average SSL\_SCORE** **value of Males** | **Difference between Average SSL\_SCORE** **values of Females and Males** | **T-Score** | | **P-Value** | |
| < 20 | 462.3939 | 482.6061 | 20.2122 | | -4.0849 | | 0.0001251 |
| 20-30 | 472.6667 | 476.0000 | 3.333 | | -0.32453 | | 0.7499 |

\*30-40: Only 6 records are present in this age group and their gender is Male and no female records are present

\*40-80: No records of this age group present in first 1000.

1. **SSL\_RANK and SEX\_CODE\_CD**
2. Without PSM (SSL\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males without PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 548.0909 | 468.9716 | 79.1193 | 1.5335 | 0.1341 |
| 20-30 | 385.6667 | 525.5625 | 139.8958 | -1.1512 | 0.2808 |

\*30-40: Only 6 records are present in this age group and their gender is Male and no female records are present

\*40-80: No records of this age group present in first 1000.

1. With PSM (SSL\_RANK and SEX\_CODE\_CD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **T-test values according to age brackets for females and males with PSM** | | | | | |
| **Age Brackets** | **Average SSL\_SCORE\_RANK value of Females** | **Average SSL\_SCORE\_RANK value of Males** | **Difference between Average SSL\_SCORE\_RANK values of Females and Males** | **T-Score** | **P-Value** |
| < 20 | 548.0909 | 244.6061 | 303.4848 | 4.3614 | 4.809e-05 |
| 20-30 | 385.6667 | 311.2222 | 74.4445 | 0.50799 | 0.6191 |

\*30-40: Only 6 records are present in this age group and their gender is Male and no female records are present

\*40-80: No records of this age group present in first 1000.

**Code and data links:**

|  |  |  |
| --- | --- | --- |
|  | **Code Git Hub Link** | **Dataset Git Hub Link** |
| **Section 1: Analysis of overall Data** |  |  |
| **Analysis of overall Data with & without PSM** | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20OVERALL%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_Full.csv |
| **Section 2: Analysis of Overall Data by Age with and without PSM** |  |  |
| <20 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%20LESS%20THAN%2020%20.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_LESS\_THAN\_20.csv |
| 20-30 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2020-30.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_20\_30.csv |
| 30-40 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2030-40.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_30\_40.csv |
| 40-50 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2040-50.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_40\_50.csv |
| 50-60 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2050-60.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_50\_60.csv |
| 60-70 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2060-70.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_60\_70.csv |
| 70-80 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20AGE%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2070-80.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_RAT\_SEX\_Code\_70\_80.csv |
| **Section 3: Analysis of first 1000 Records** |  |  |
| **Analysis of first 1000 Records with and without PSM** | <https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20FIRST%201000%20%20SSL_SCORE%20SSL_SCORE_RANK%20and%20SEX_CODE_CD%20.ipynb> | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_FIRST\_1000\_SEX\_Code.csv |
| **Section 4: Analysis of first 1000 Records by age with and without PSM** |  |  |
| <20 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20FIRST%201000%20by%20Age%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%20Less%20than%2020.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_FIRST\_1000\_SEX\_Code\_Less\_than\_20.csv |
| 20-30 | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/0711%20FIRST%201000%20by%20Age%20SSL\_SCORE%20SSL\_SCORE\_RANK%20and%20SEX\_CODE\_CD%2020-30.ipynb | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_FIRST\_1000\_SEX\_Code\_20\_30.csv |
| 30-40 | NA | https://github.com/kshitijkhurana3010/Algorithmic-Bias-Project/blob/master/SSL\_FIRST\_1000\_SEX\_Code\_30\_40.csv |