**EDA – Numerical Features**

|  |  |
| --- | --- |
| *Age at Injury* | * Dropped values <16 * Filled NAs with the average (because the distribution is simetric) * Handled outliers with zscore |
| *Average Weekly Wage* | * No missing values * Handled outliers with zscore * KNNImputer |
| *Birth Year* | * Dropped values < 1903 * Filled NAs with Accident Date – Age at Injury * Handled outliers with zscore |
| *Claim Identifier* | * No missing values * Handled outliers with zscore |
| *Industry Code* | * No missing values * Handled outliers with zscore * KNNImputer |
| *IME-4 Count* | * Filled NAs with zero * Handled outliers replacing all values > 15 with 15 |
| *WCIO Cause of Injury Code* | * Filled NAs with zero (resolver) * Handled outliers with zscore |
| *WCIO Nature of Injury Code* | * Filled NAs with zero (resolver) * Handled outliers with zscore |
| *WCIO Part of Body Code* | * Filled NAs with zero (resolver) * Handled outliers with zscore |
| *Number of Dependents* | * Filled NAs with zero (resolver) * Handled outliers with zscore |

**Novas:**

|  |  |
| --- | --- |
| *Accident Date* | * Dropped NAs |
| *Years Past Accident* | * Accident Date – reference\_date |

**Missing values before:**

* Accident Date: 0.64%
* Age at Injury: 0.94%
* Average Weekly Wage: 4.99%
* Birth Year: 0.38%
* Industry Code: 1.73%
* Years Past Accident: 0.64%

**Missing values after:**

* Average Weekly Wage: 4.99%
* Industry Code: 1.46%