**Hackathon : Data Science & Analytics**

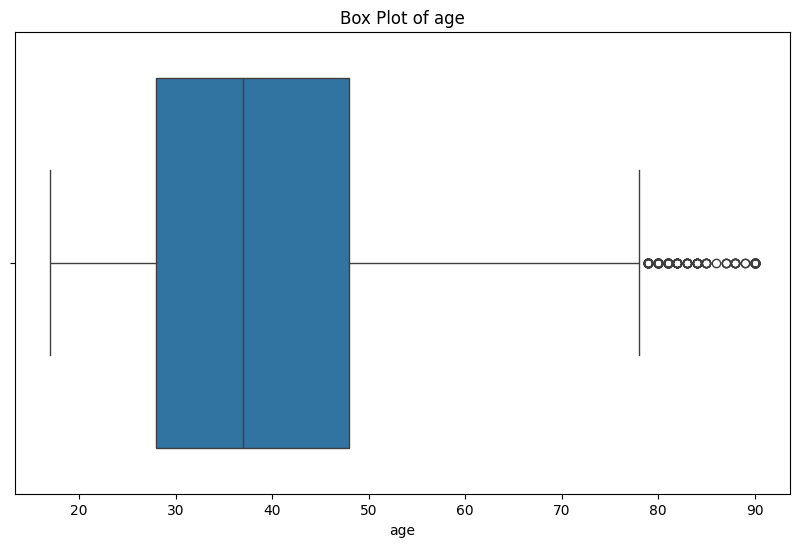
**Challenge: From Data to Insights**

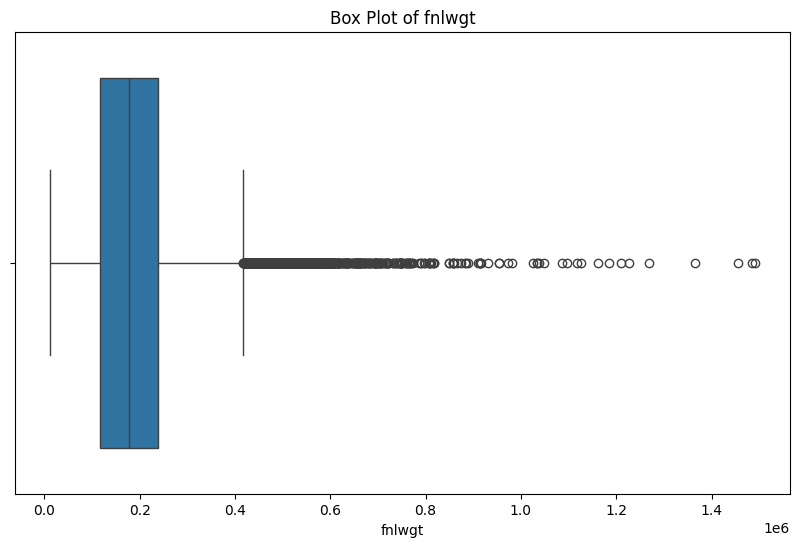
1. **Automated Data Cleaning and Preprocessing Challenge**

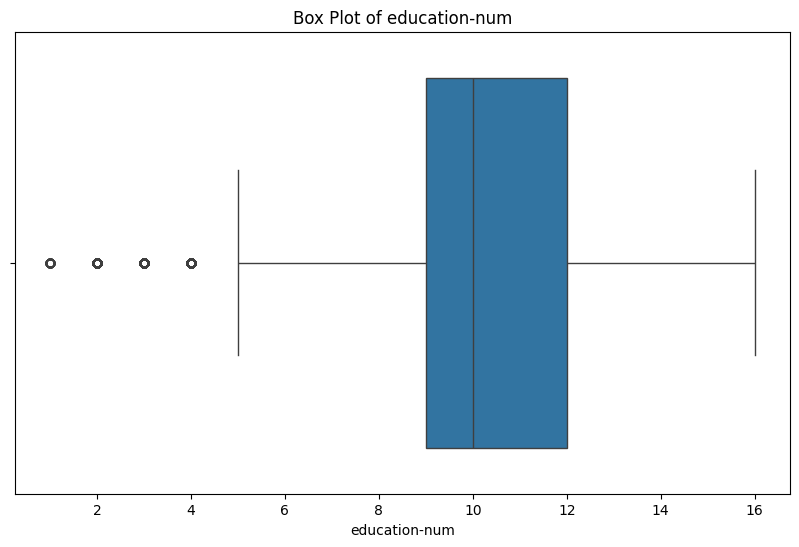
Dataset: <https://archive.ics.uci.edu/dataset/2/adult>

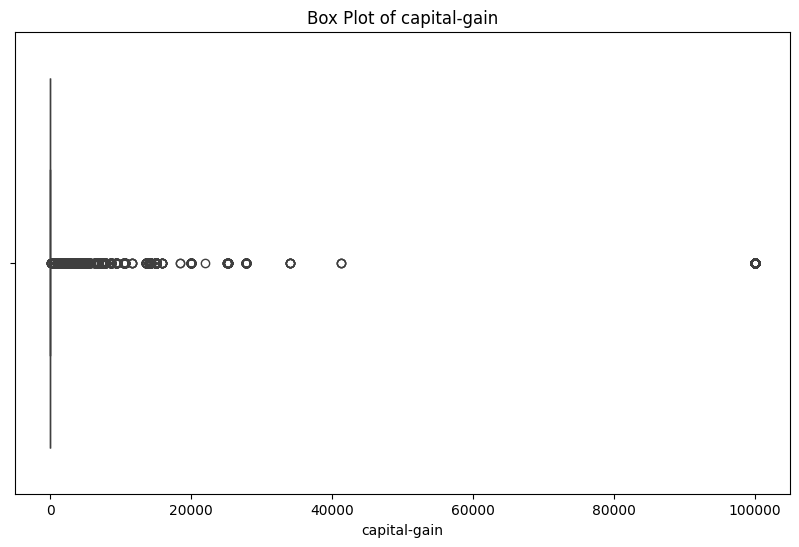
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **name** | **role** | **type** | **demographic** | **description** | **units** | **missing\_values** |
| age | Feature | Integer | Age | N/A |  | no |
| workclass | Feature | Categorical | Income | Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked | | yes |
| fnlwgt | Feature | Integer |  | None |  | no |
| education | Feature | Categorical | Education Level | Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool | | no |
| education-num | Feature | Integer | Education Level | None |  | no |
| marital-status | Feature | Categorical | Other | Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse | | no |
| occupation | Feature | Categorical | Other | Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces | | yes |
| relationship | Feature | Categorical | Other | Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried | | no |
| race | Feature | Categorical | Race | White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black | | no |
| sex | Feature | Binary | Sex | Female, Male. |  | no |
| capital-gain | Feature | Integer |  | None |  | no |
| capital-loss | Feature | Integer |  | None |  | no |
| hours-per-week | Feature | Integer |  | None |  | no |
| native-country | Feature | Categorical | Other | United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands | | yes |
| income | Target | Binary | Income | >50K, <=50K. |  | no |

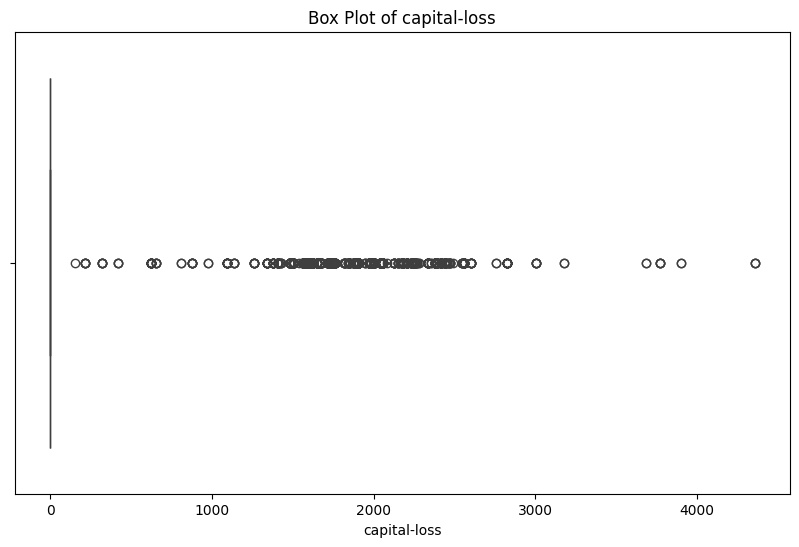
**BOX PLOT OF DATA**

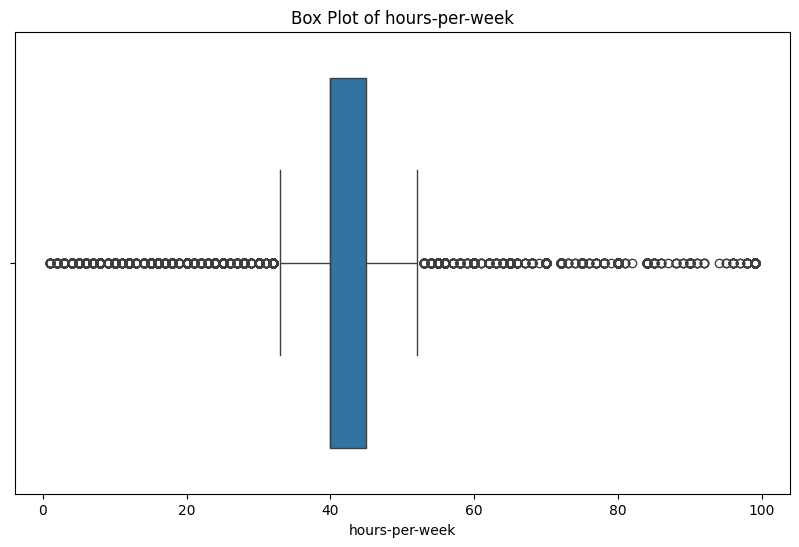
****

****

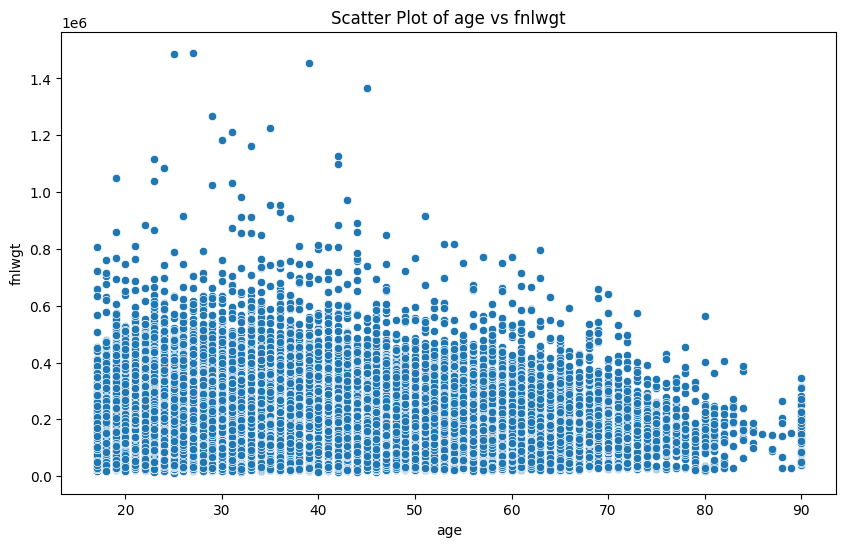
****

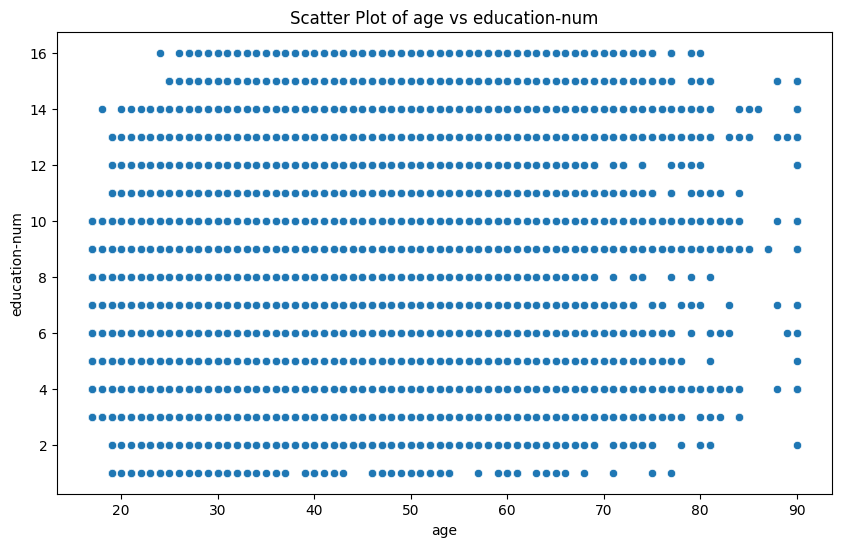
****

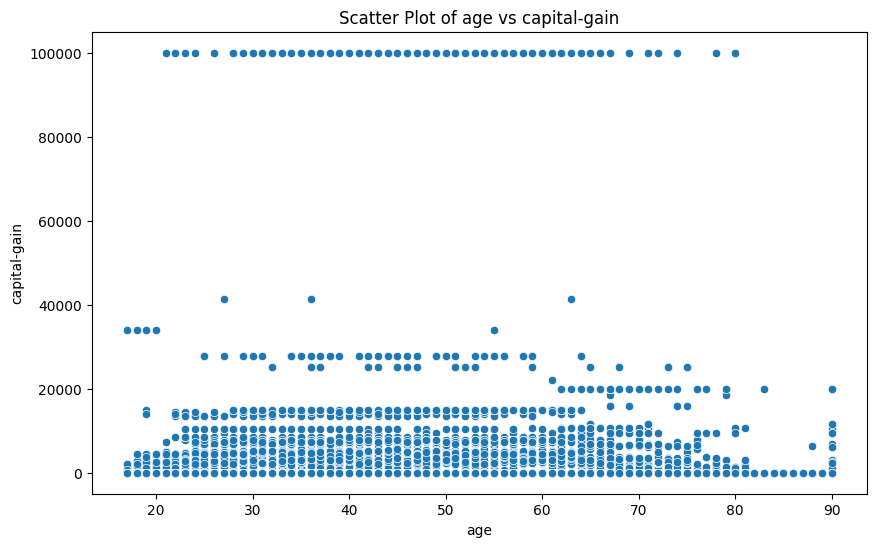
****

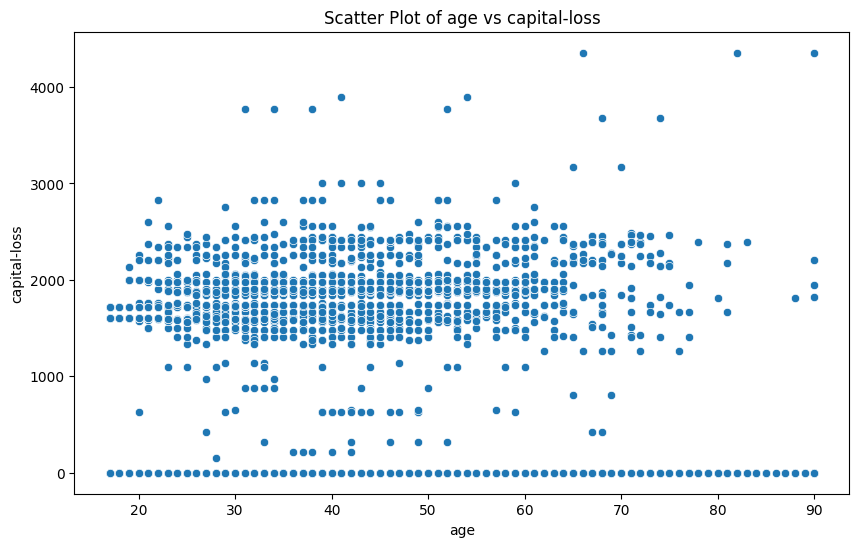


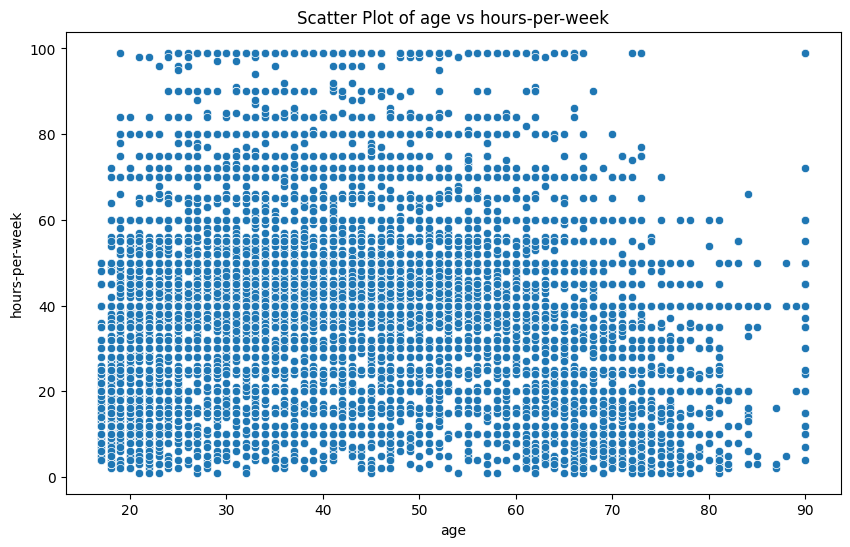
**SCATTER PLOTS FOR PAIRWISE COMPARISON**

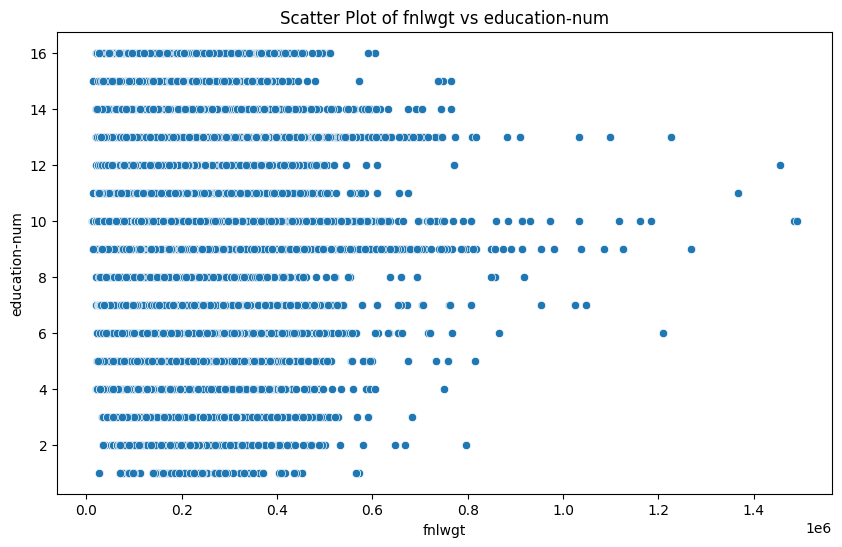
****

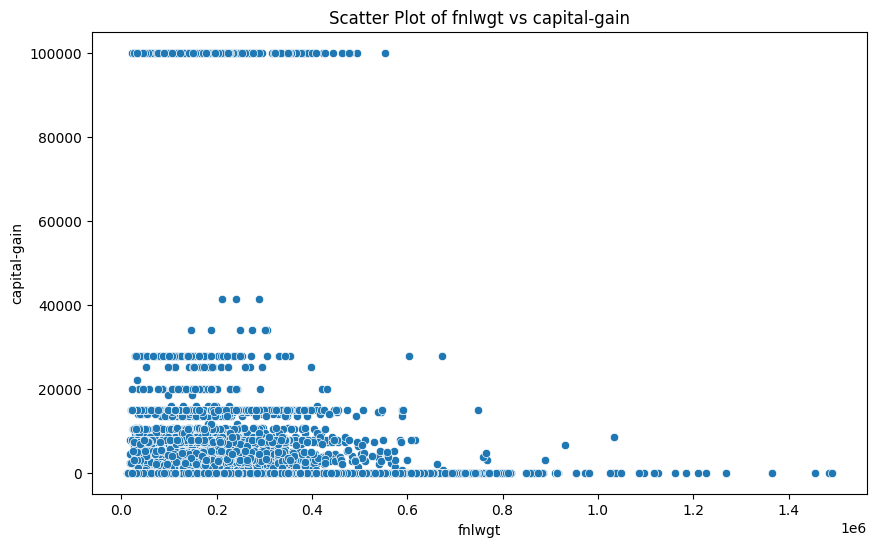
****

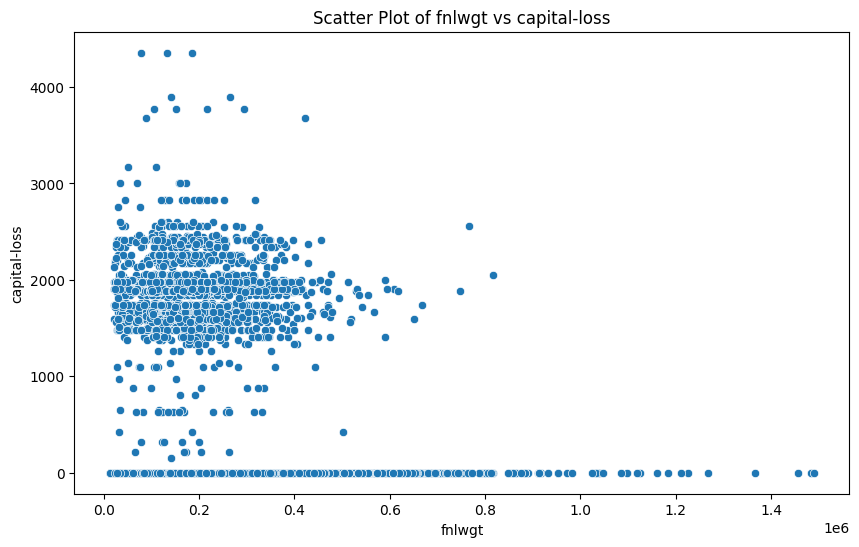
****

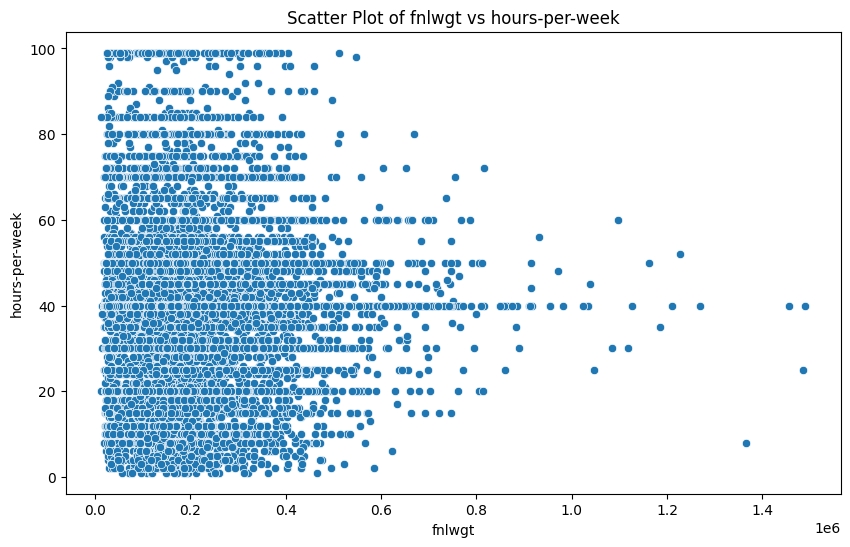
****

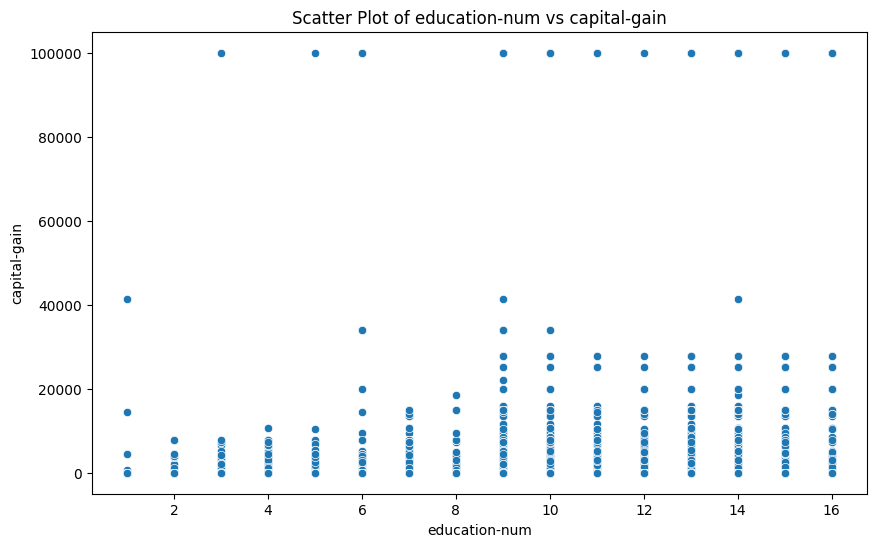
****

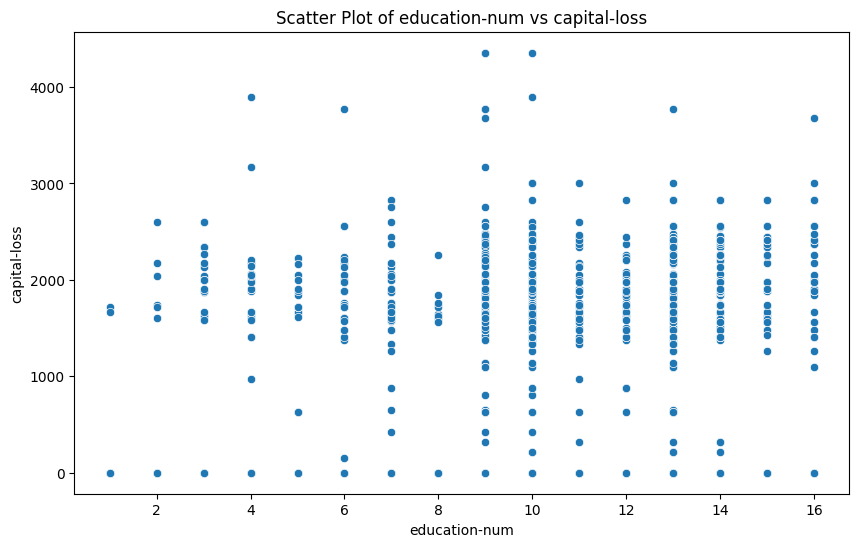
****

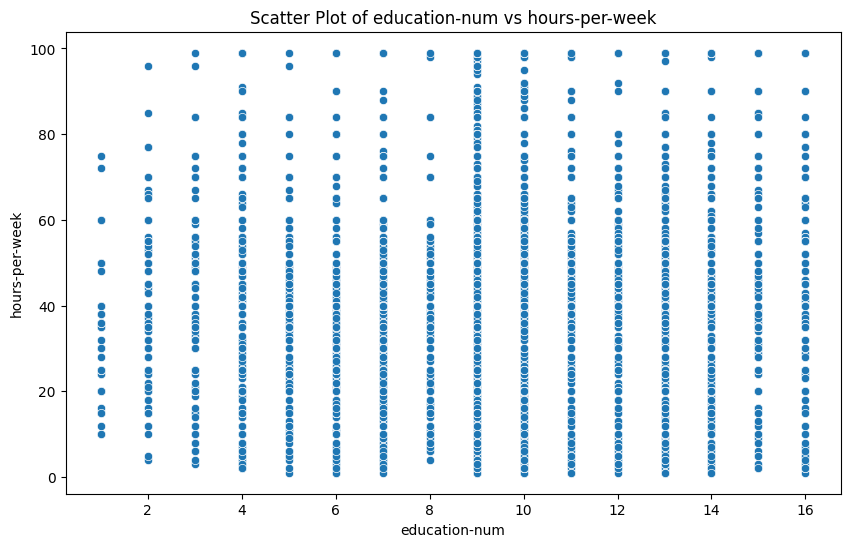
****

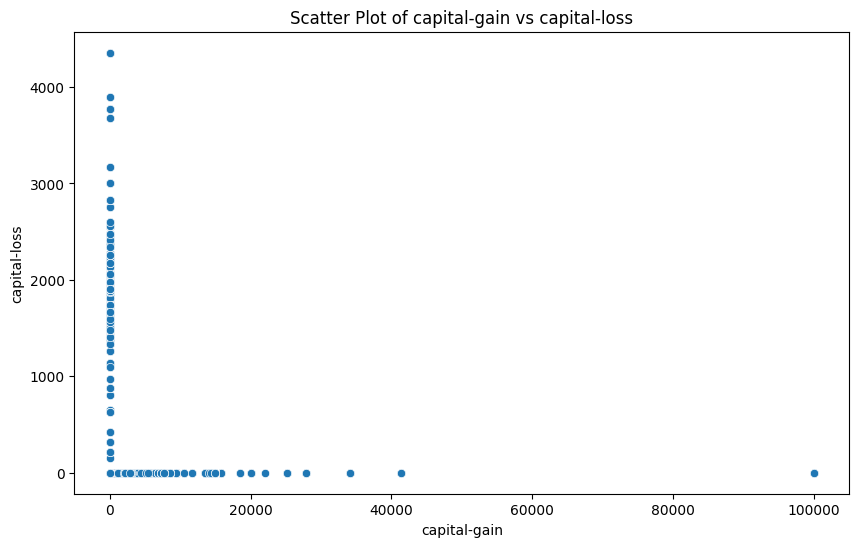
****

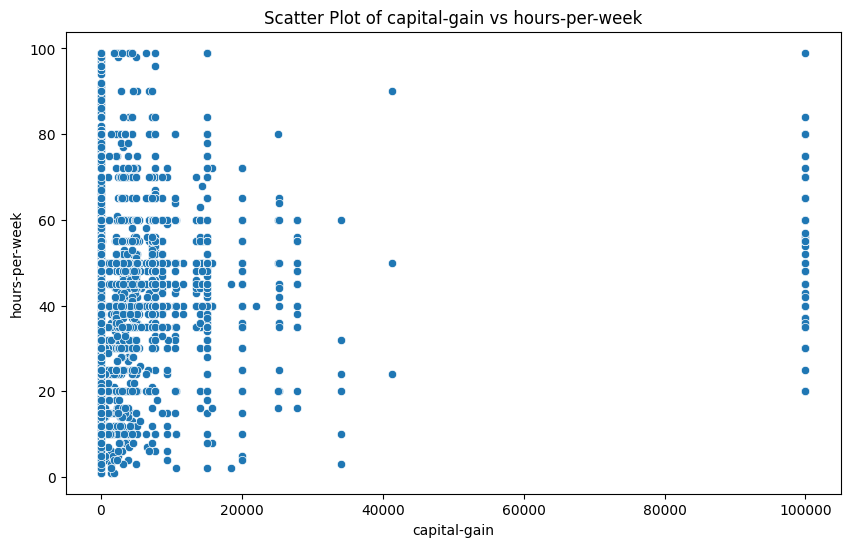
****

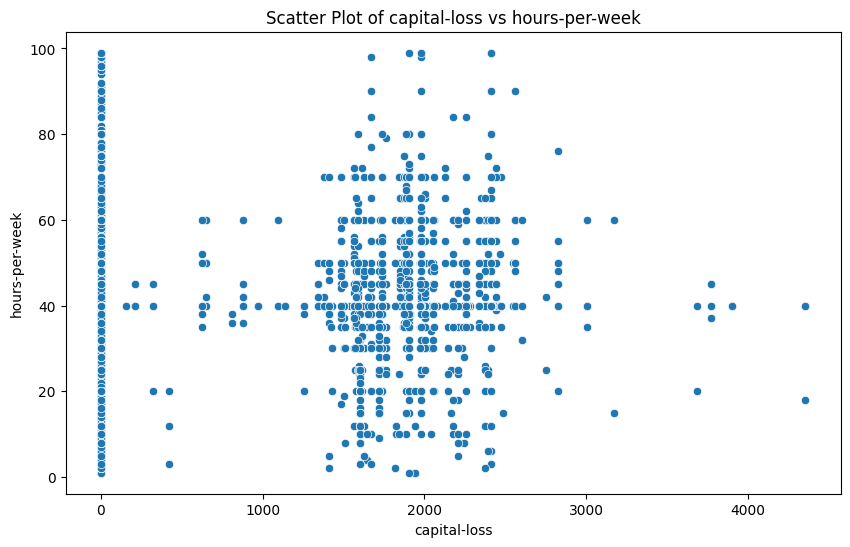
****

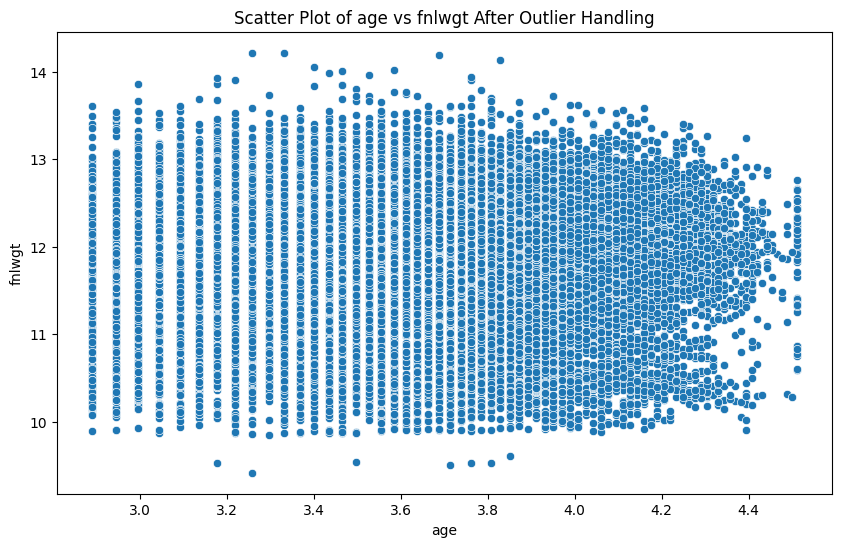
****

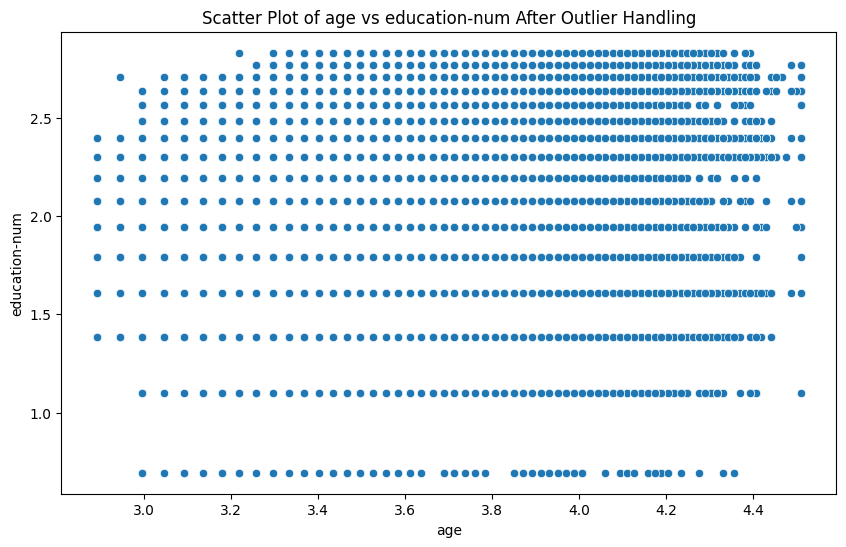
****

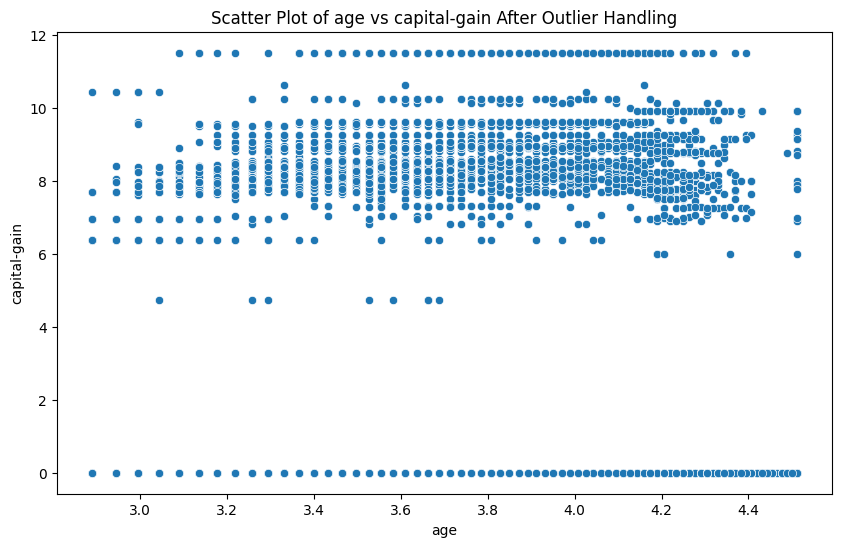
****

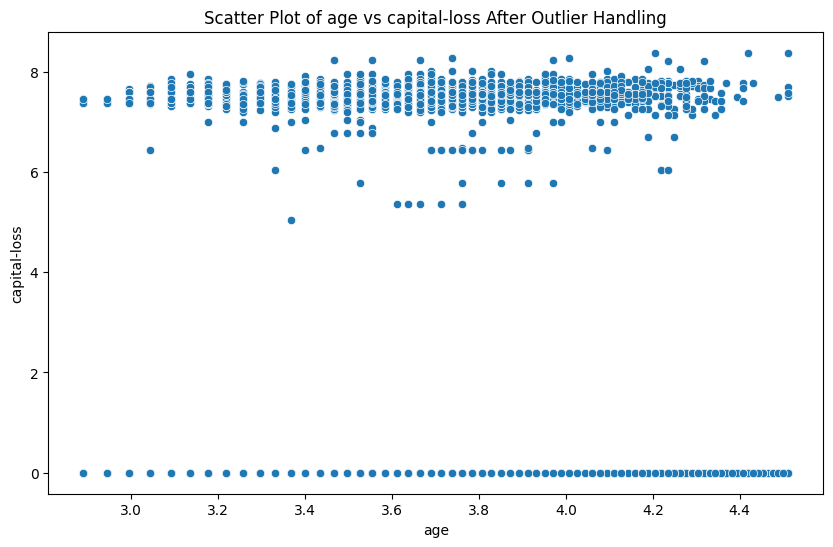
****

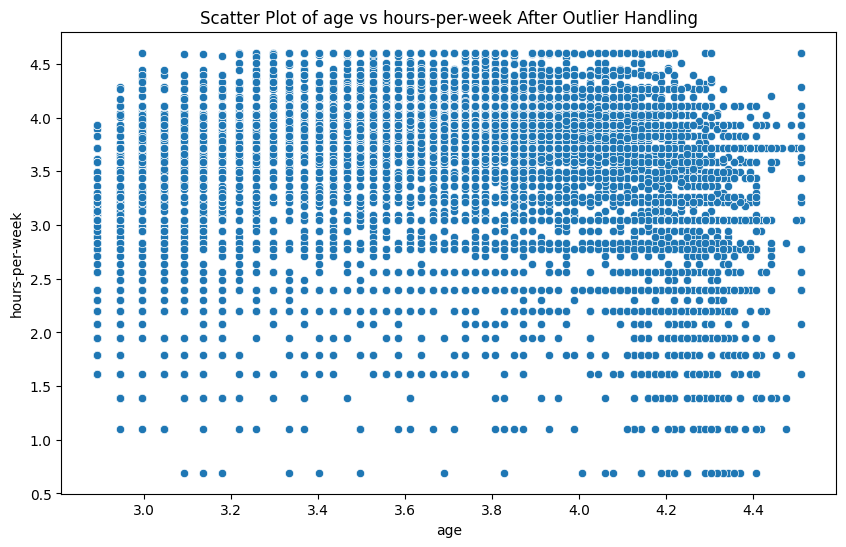
****

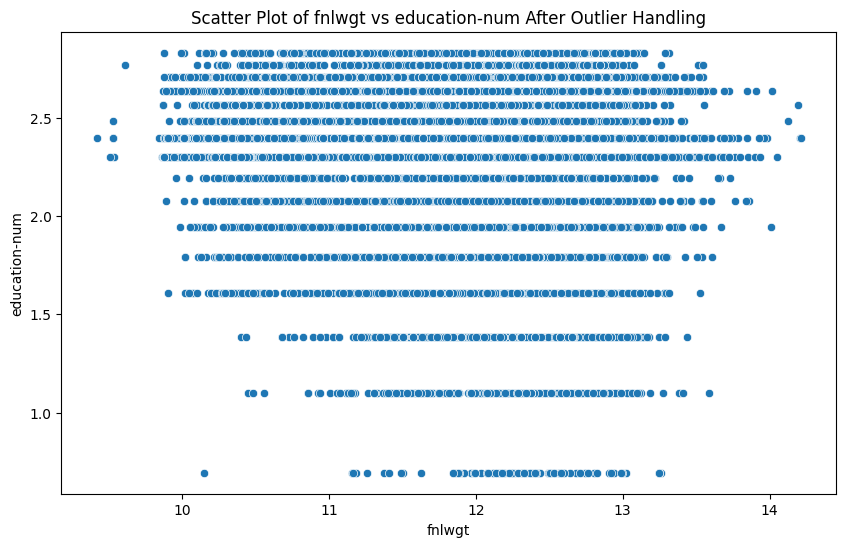
****

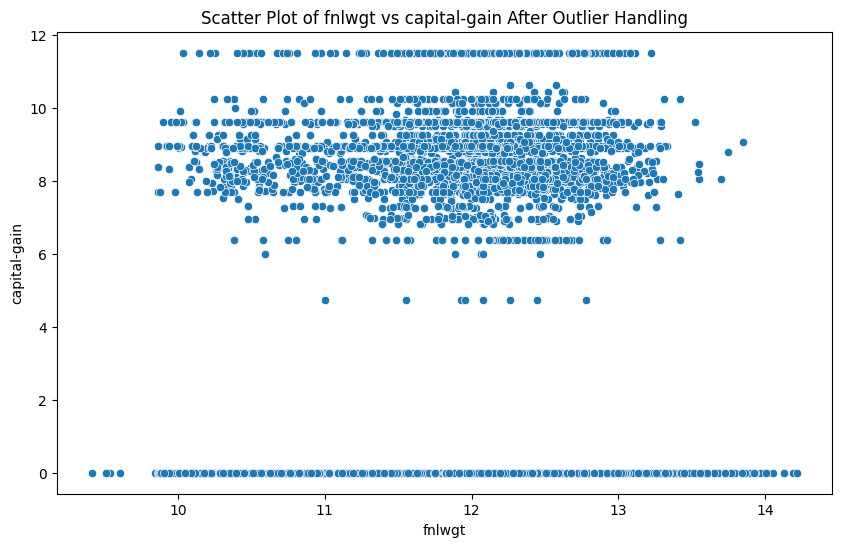
****

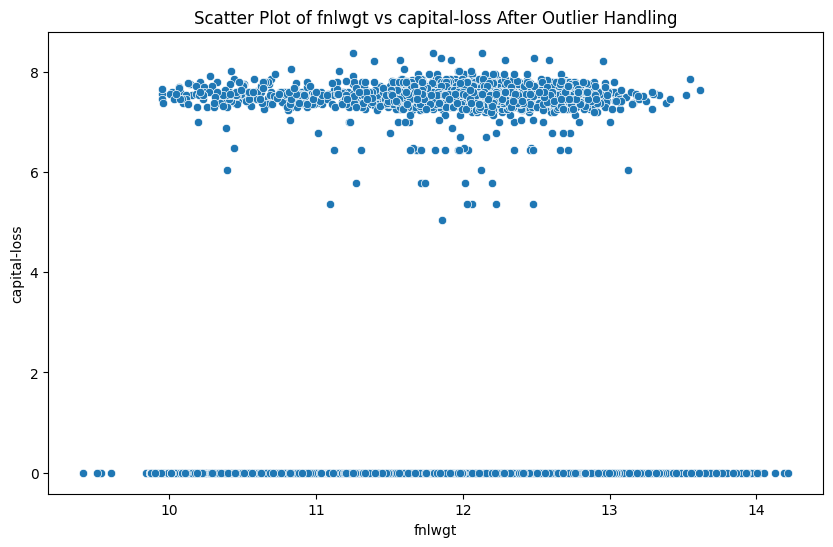
****

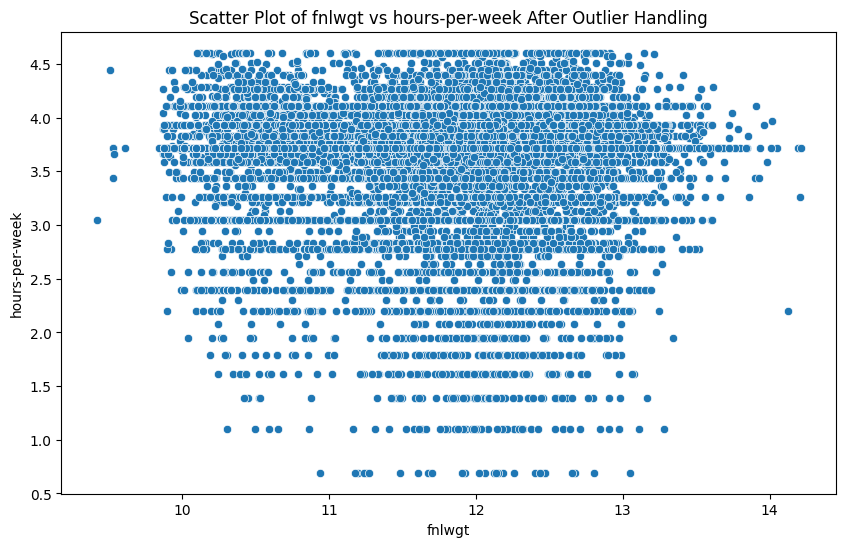
****

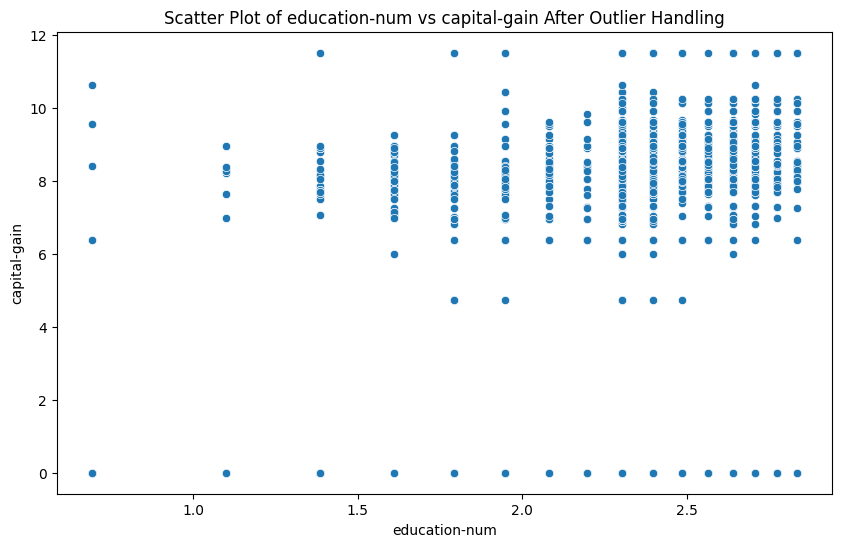
****

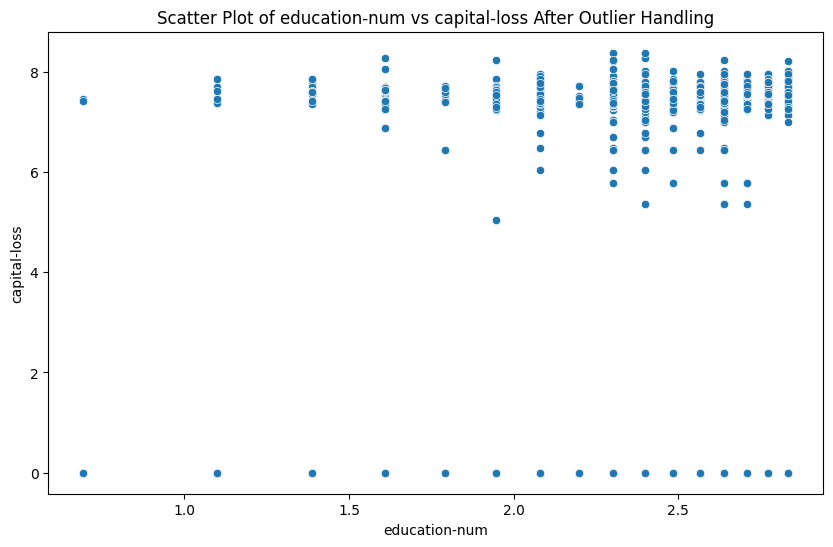
****

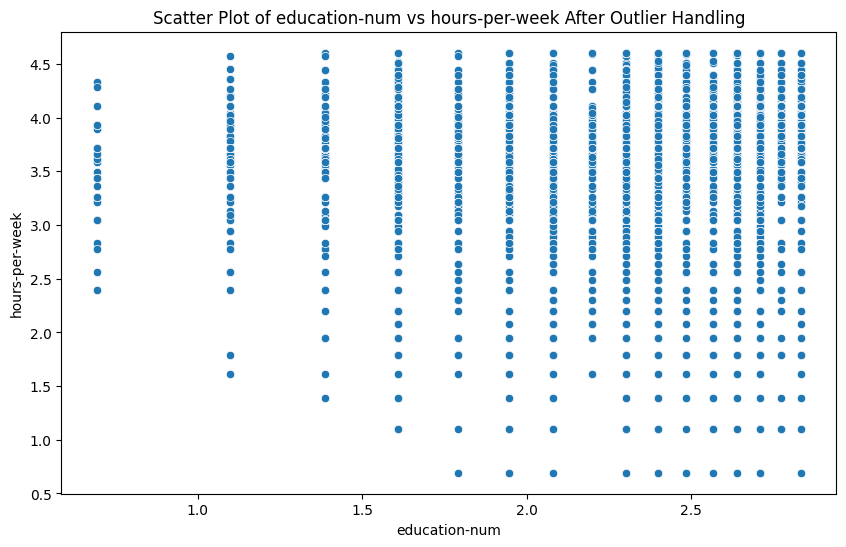
****

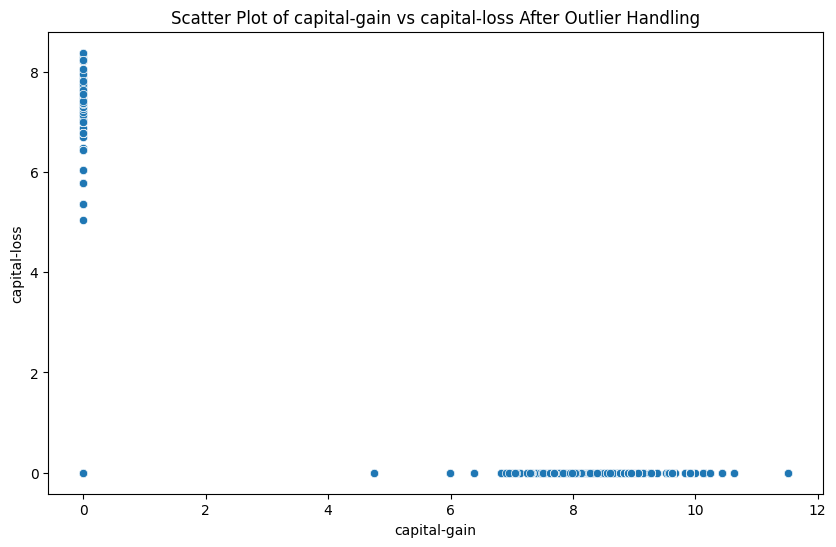
****

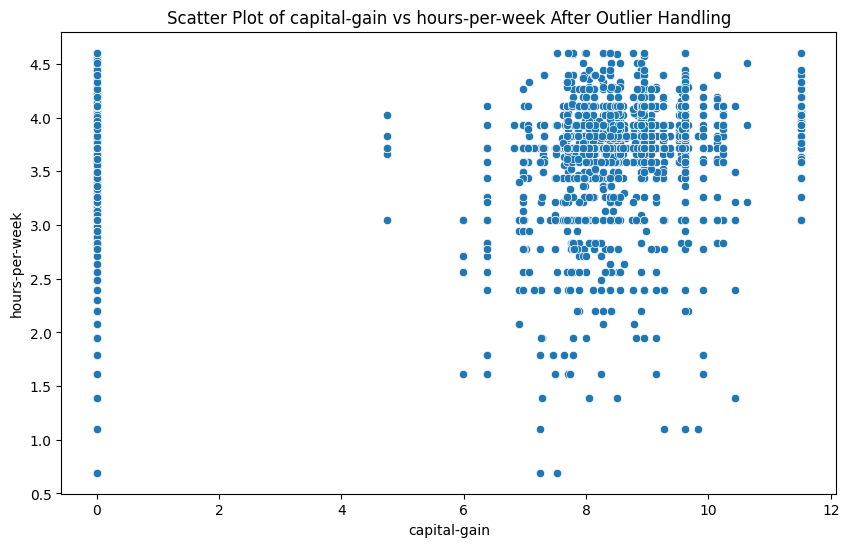
****

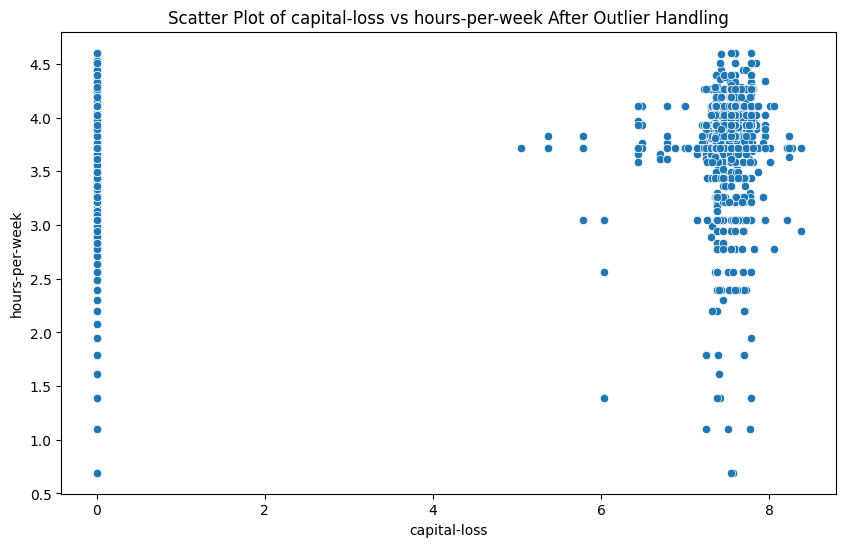
****

****

****

****

****

****

**Accuracy with all features:** 0.4385300440167878

**Classification Report with all features:**

**precision recall f1-score support**

<=50K 0.63 0.52 0.57 4936

<=50K. 0.34 0.31 0.33 2478

>50K 0.35 0.42 0.38 1562

>50K. 0.18 0.35 0.24 793

**accuracy**  0.44 9769

**macro avg** 0.38 0.40 0.38 9769

**weighted avg** 0.48 0.44 0.45 9769