 **Indiana University, Bloomington**

**DSCI-D 590 Biomedical Data Science in Practice**

**Final Report**

**Team Members:**

Priyanka Prem Kumar

[prpremk@iu.edu](mailto:prpremk@iu.edu)

Madhuri Patibandla

[mpatiba@iu.edu](mailto:mpatiba@iu.edu)

Madhumathi Sekar

[msekar@iu.edu](mailto:msekar@iu.edu)

**Investigating clinical risk factors for ICI-Associated Acute Kidney injury using real-world data and Causal Machine Learning**

**ABSTRACT**

Immune checkpoint inhibitors (ICIs) have transformed the landscape of cancer therapy by enhancing the immune system's ability to target and eliminate malignant cells. However, their use is associated with immune-related adverse events, including ICI-associated acute kidney injury (ICI-AKI), which can lead to significant morbidity and treatment discontinuation. Understanding how clinical, demographic, phenotypic, visit-related, social, and insurance factors contribute to ICI-AKI risk is essential for improving patient stratification and advancing precision oncology.

This study aims to move beyond traditional association-based methods to uncover causal relationships between key risk factors and the development of ICI-AKI. Leveraging the nationally representative All of Us research cohort, which links rich electronic health record (EHR) data with self-reported social determinants of health, we analyze a cohort of ICI-treated patients to investigate patterns of kidney injury and associated risk profiles.

We begin with exploratory analysis of ICI exposure, AKI incidence, and clinical covariates such as age, sex, race, comorbidities, and insurance coverage. Traditional approaches including logistic regression and Kaplan-Meier survival analysis are used to estimate associations and timing of ICI-AKI events. These results are then compared with those obtained through advanced causal inference techniques, including Targeted Maximum Likelihood Estimation (TMLE) and Causal Forests, which aim to estimate heterogeneous treatment effects and account for complex confounding.

By comparing conventional statistical models with modern causal machine learning approaches, this project offers a comprehensive evaluation of ICI-AKI risk factors, with careful attention to underlying assumptions and confounders. The integration of causal inference and machine learning not only strengthens the validity of our findings but also enhances their clinical utility. Ultimately, our results have the potential to inform biomarker discovery, guide personalized treatment decisions, and support proactive mitigation strategies for renal toxicity in immunotherapy-treated populations.

**INTRODUCTION**

Immune checkpoint inhibitors (ICIs) have revolutionized cancer treatment by enabling durable responses across a variety of malignancies such as melanoma, non-small cell lung cancer, and renal cell carcinoma. These therapies act by blocking inhibitory pathways such as CTLA-4, PD-1, or PD-L1, thereby reactivating T-cell mediated anti-tumour immunity and improving survival outcomes in diverse patient populations. However, their clinical use is often complicated by immune-related adverse events (irAEs), including immune checkpoint inhibitor-associated acute kidney injury. ICI-AKI poses significant challenges, leading to treatment interruptions, decreased quality of life, and increased healthcare resource utilization (Cortazar et al., 2016).

Previous research has identified key clinical risk factors for ICI-AKI, such as age, baseline renal function, and concurrent use of nephrotoxic agents (Seethapathy et al., 2019). Despite these advances, understanding of the causal mechanisms and patient-specific susceptibility remains limited, particularly given the complexity of immune-mediated toxicity and variability in patient responses. Traditional association studies including logistic regression and survival analysis are valuable but may not adequately address confounding and selection biases inherent in observational data.

Our prior analyses using Logistic regression, Kaplan-Meier survival methods, and advanced causal inference techniques such as Targeted Maximum Likelihood Estimation (TMLE), Propensity Score Weighting, and Inverse Probability Weighting (IPW) laid the groundwork by identifying potential risk factors and estimating their associations with ICI-AKI. Building on this, the current study leverages cutting-edge machine learning approaches specifically Causal Forests to further improve estimation of heterogeneous treatment effects and to better account for complex confounding structures.

The All of Us (AoU) Research Program offers a unique, ethnically diverse cohort of over 400,000 participants with linked longitudinal electronic health records (EHRs) and socio-demographic data, providing a rich resource to examine clinical and social determinants of ICI-AKI (AoU Program Investigators, 2019). Utilizing this multidimensional dataset, we compare traditional statistical models with causal machine learning frameworks to assess the robustness, validity, and interpretability of identified risk factors. This study emphasizes a thoughtful comparison between traditional statistical approaches and modern causal machine learning methods, with careful consideration of underlying assumptions, confounding structures, and the strengths and limitations inherent to each framework.

By integrating causal inference with machine learning methods in the context of real-world data, this work aims to enhance precision oncology through improved risk stratification and targeted prevention strategies for ICI-AKI. Our findings will contribute to better clinical decision-making and individualized patient management in immunotherapy.

**Team Members Contributions**

|  |  |
| --- | --- |
| Priyanka Prem Kumar | Standard Logistic regression,  Feature Importance - Odd Ratios  Logistic regression with Lasso Regularization  ROC AUC curve and Cross Validation |
| Madhumathi Sekar | Kaplan-Meier Survival Analysis  Cox Proportional Hazards model |
| Madhuri Patibandla | Targeted Maximum Likelihood Estimation,  Causal Forest DML  Propensity Model and with Tuning,  Comparing Traditional vs Causal ML Methods |

**METHODOLOGY**

This study employs a dual-method approach, integrating traditional statistical modelling with modern causal machine learning techniques to identify and estimate the potential causal impact of phenotypic traits, visit-related factors, comorbidities, insurance data and social determinants of health on the risk of ICI-AKI. The analysis leverages data from the All of Us (AoU) Research Program, a nationally representative and ethnically diverse cohort that combines electronic health records (EHRs), detailed visit histories, phenotype data, and self-reported demographic and social determinants of health information.

**1. Cohort Construction**

An ICI-treated cohort was constructed by identifying participants who received at least one immune checkpoint inhibitor therapy. Within this cohort, individuals were categorized into AKI and non-AKI groups based on the presence or absence of acute kidney injury diagnosis codes. To reduce immortal time bias and ensure comparability across individuals, observation windows were aligned relative to the initiation of ICI treatment.

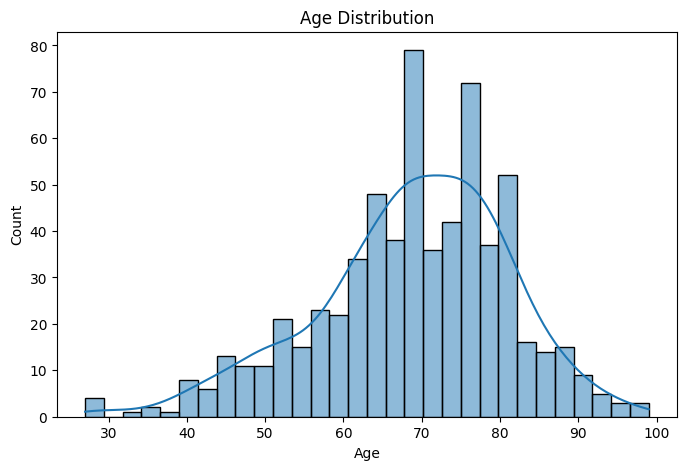
**2. Data Setup**

Phenotype, visit-level, comorbidity, insurance, and social determinants of health (SDOH) data were securely accessed and downloaded from a protected cloud storage environment into the local analysis workspace. Visit-level records were aggregated to the person level by summarizing key metrics such as the total number of visits and inpatient encounters per individual. Comorbidity data were similarly aggregated by computing individual-level counts or flags for the presence of specific conditions. Insurance and SDOH data were processed to reflect person-level attributes, such as primary insurance type or socioeconomic status indicators. All datasets were standardized and merged using unique patient identifiers, resulting in a comprehensive, patient-level analytic dataset suitable for downstream causal and statistical analyses.

**3. Data Visualization**

To explore patterns, identify potential confounders, and assess the distribution of key variables within the dataset, we conducted a series of descriptive and inferential visualizations. These visualizations supported hypothesis generation and informed subsequent modelling strategies.

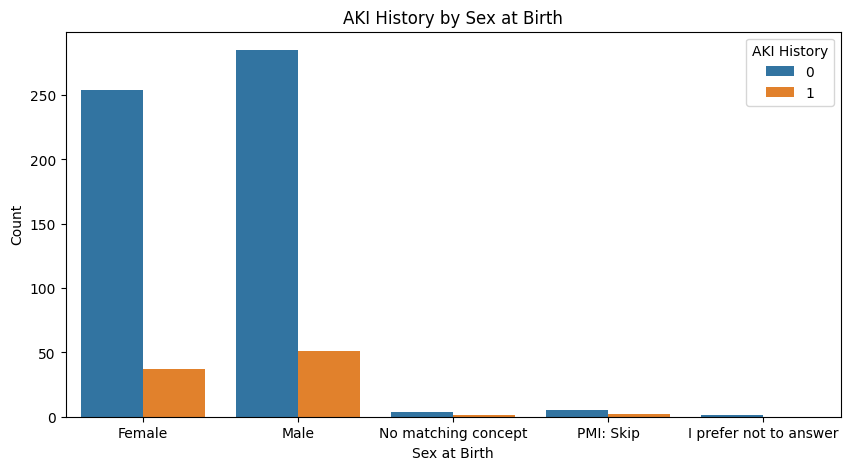
3.1. Demographic Distributions



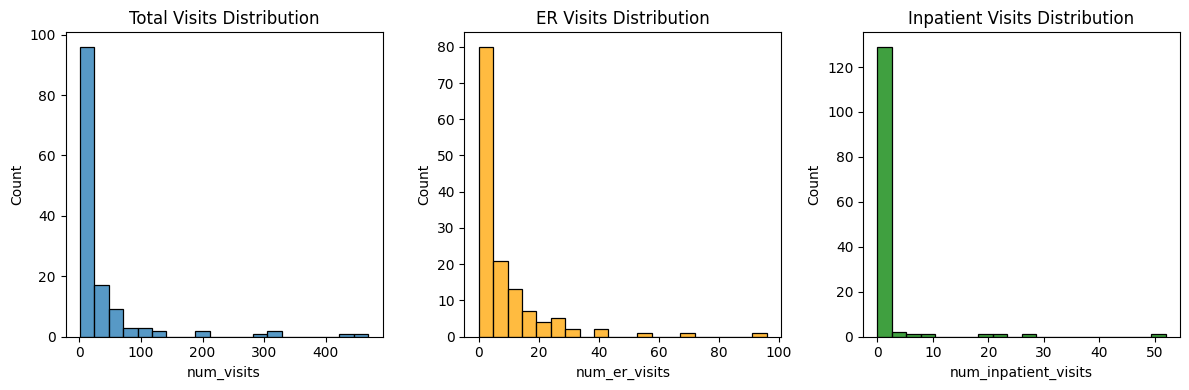
We generated histograms and count plots to characterize the demographic makeup of the cohort. The age distribution was visualized using a histogram with kernel density estimation (KDE), highlighting a skew toward older individuals.

3.2. AKI and AKI\_ICI Event Analysis

Bar plots were employed to illustrate the frequency of AKI history and AKI\_ICI events, stratified by sex at birth. These visualizations revealed differences in AKI prevalence across demographic groups and aided in identifying high-risk subpopulations.

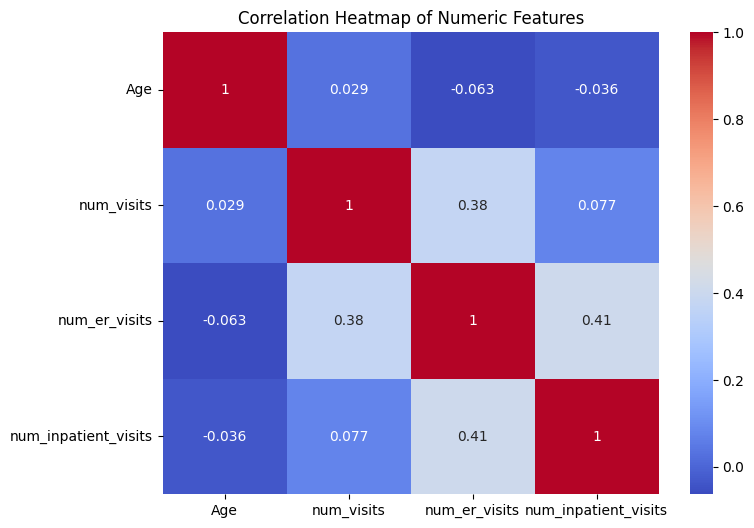


3.3. Healthcare Utilization Patterns

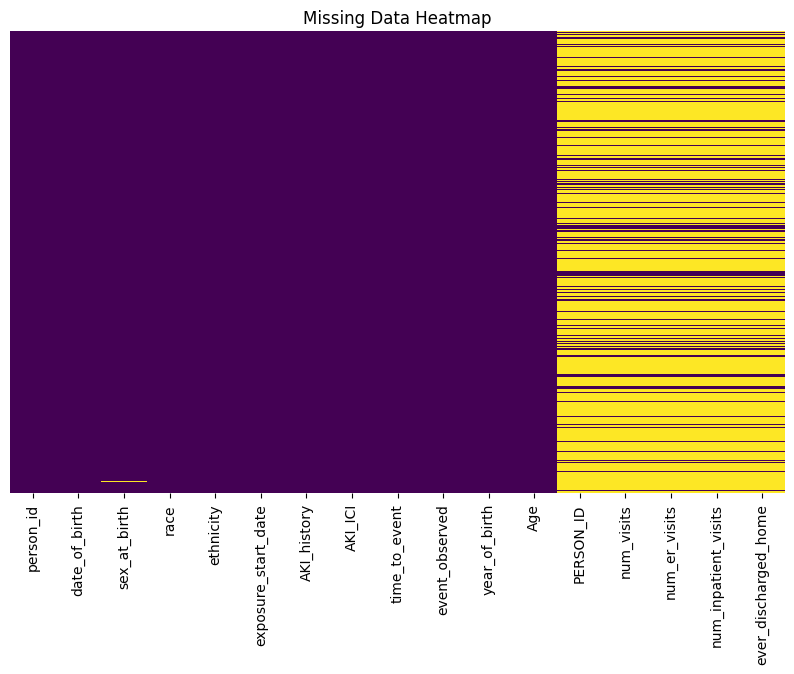


Distributions of total, emergency room (ER), and inpatient visit counts were visualized using histograms. These plots revealed skewed usage patterns and highlighted subsets of high-utilization patients.

3.4. Correlation Analysis

A heatmap of Pearson correlations among numeric features (e.g., age, number of visits) was generated to identify linear relationships and potential multicollinearity.

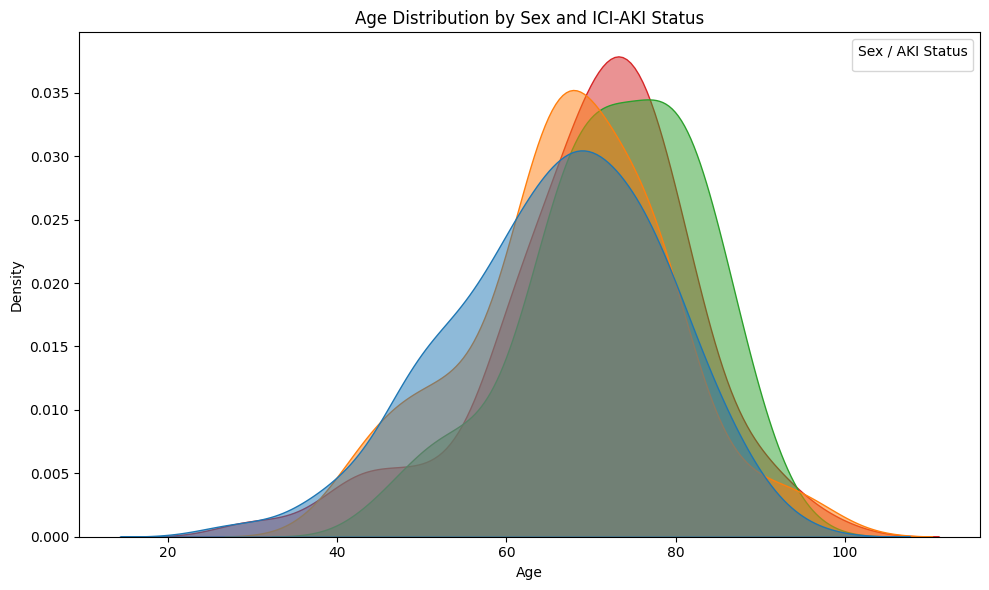
3.5. Missingness Exploration



A heatmap was created to visualize missing data across all variables. This allowed assessment of missingness patterns and informed decisions about imputation or exclusion in downstream analysis.

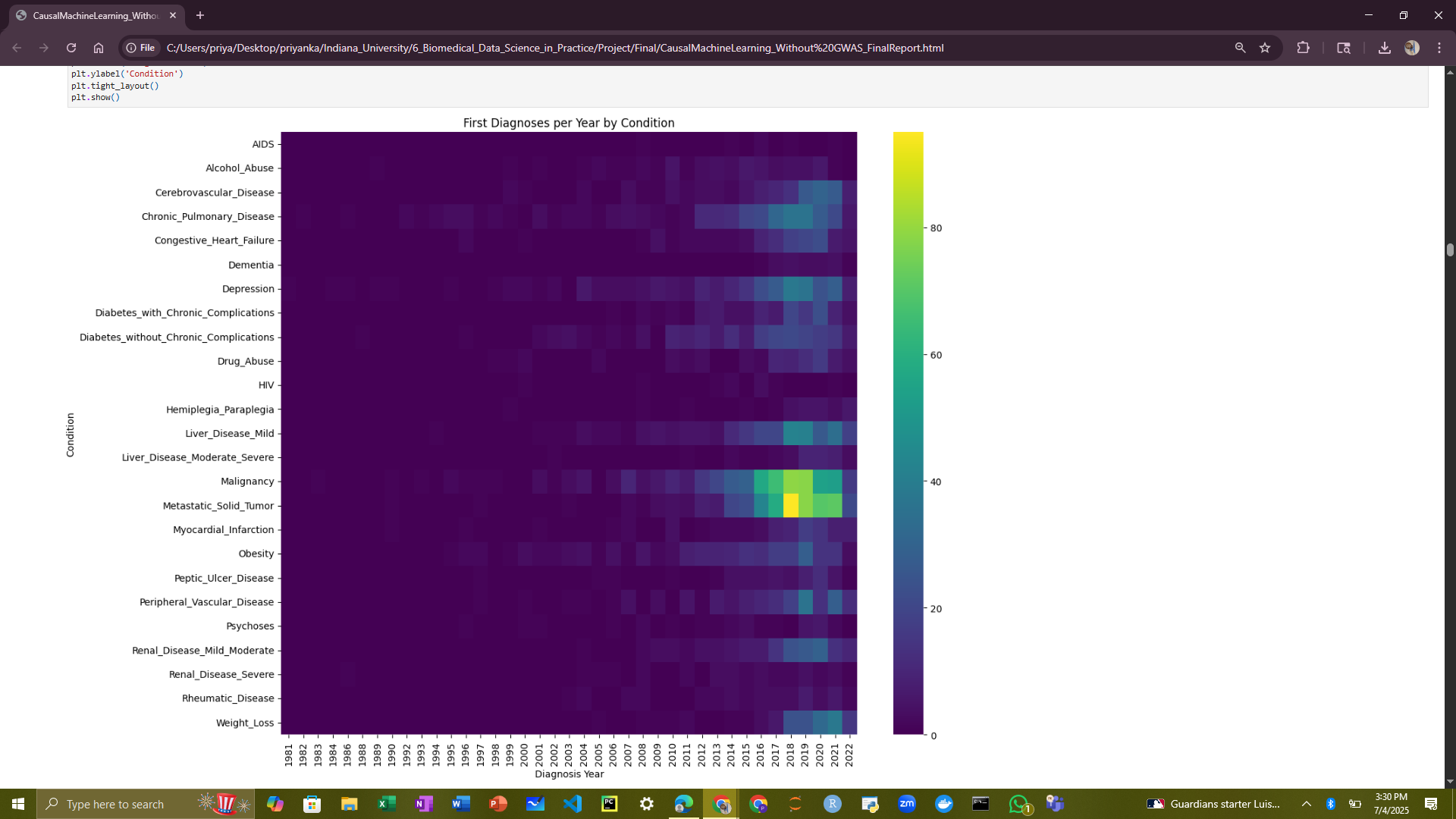
3.6. Stratified KDE Plots

To examine how age distributions varied by AKI\_ICI status within sexes, stratified kernel density plots were generated using custom subplots for male and female participants. These plots enabled side-by-side comparison of the age-related risk distribution between sexes.



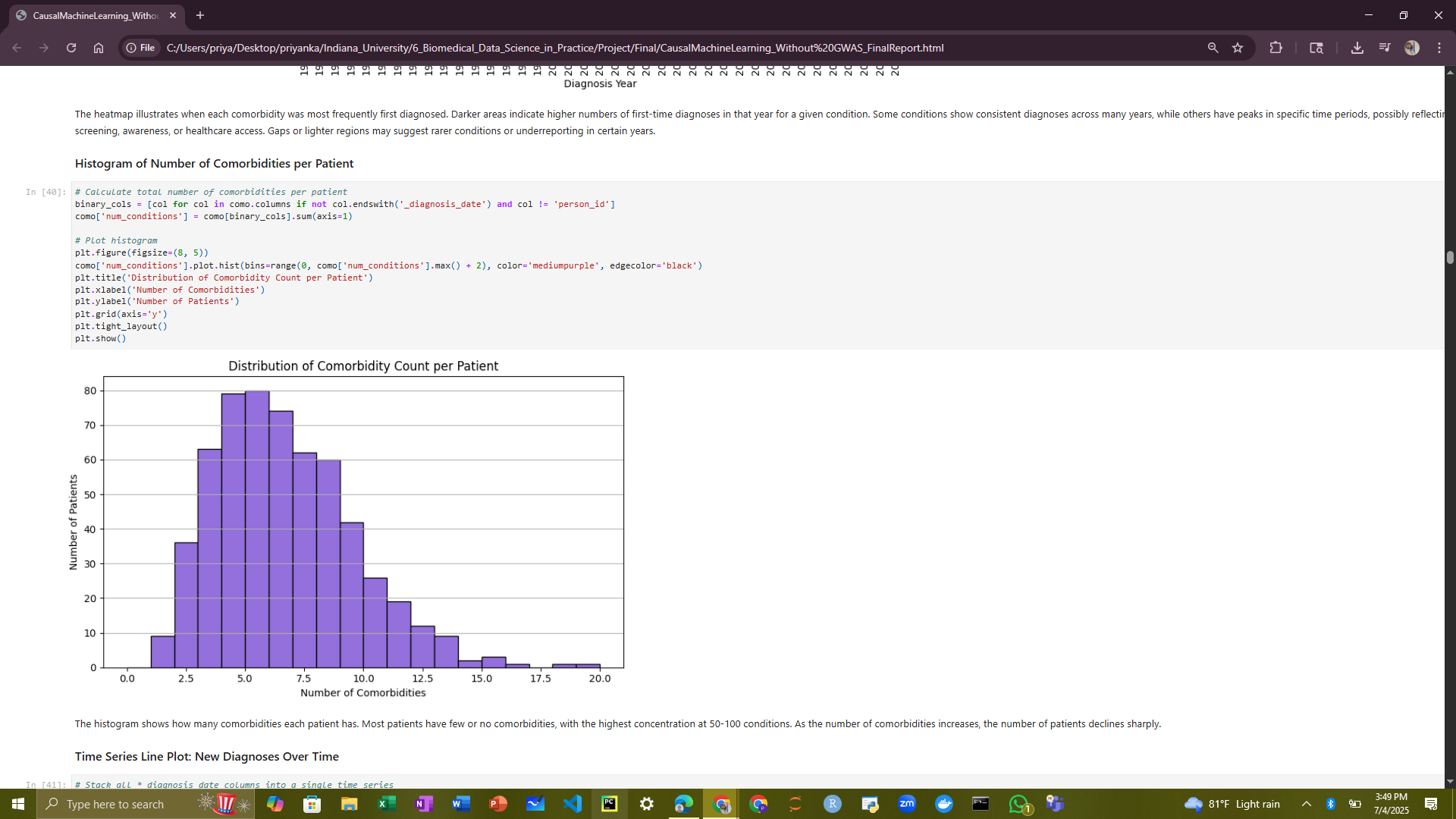
3.7 Timeline Heatmap of First Diagnosis by Condition

The heatmap illustrates when each comorbidity was most frequently first diagnosed. Darker areas indicate lower numbers of first-time diagnoses in that year for a given condition. Some conditions show consistent diagnoses across many years, while others have peaks in specific time periods, possibly reflecting shifts in screening, awareness, or healthcare access. Gaps or lighter regions may suggest rarer conditions or underreporting in certain years.

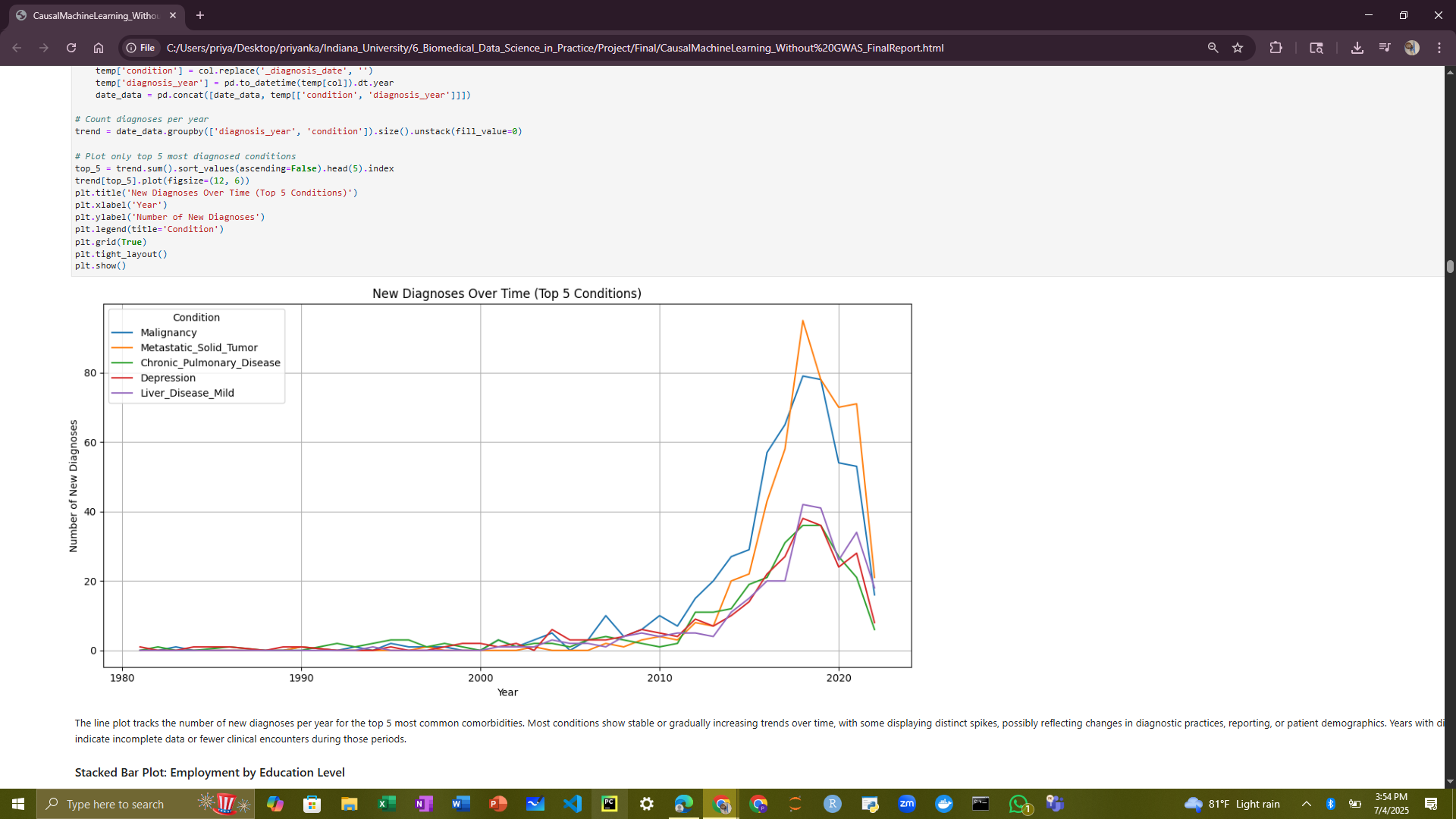


3.8 Histogram of Number of Comorbidities per Patient

The histogram shows how many comorbidities each patient has. Most patients have few or no comorbidities, with the highest concentration at 4-6 conditions. As the number of comorbidities increases, the number of patients declines sharply.

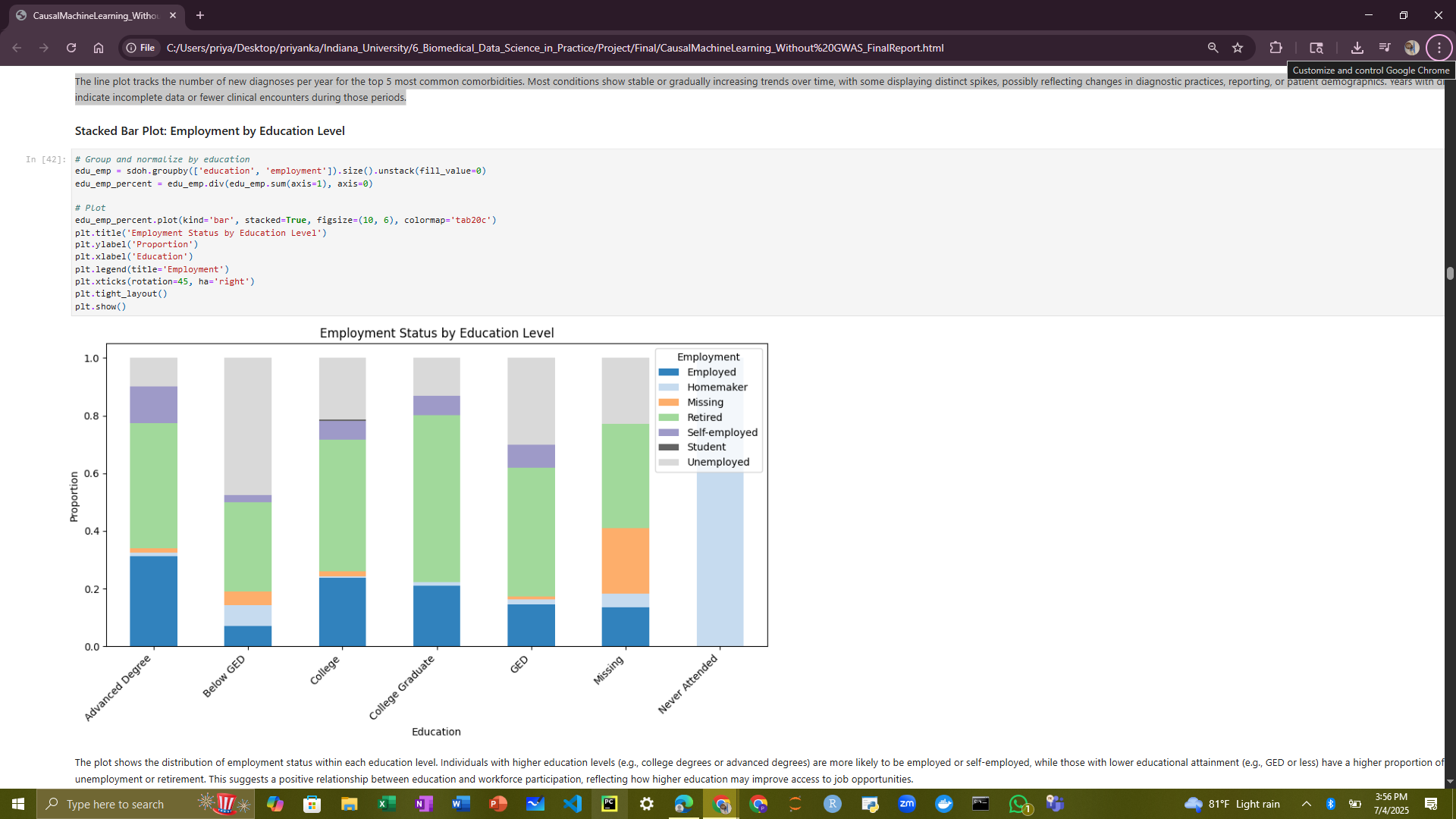


3.9 Time Series Line Plot: New Diagnoses Over Time



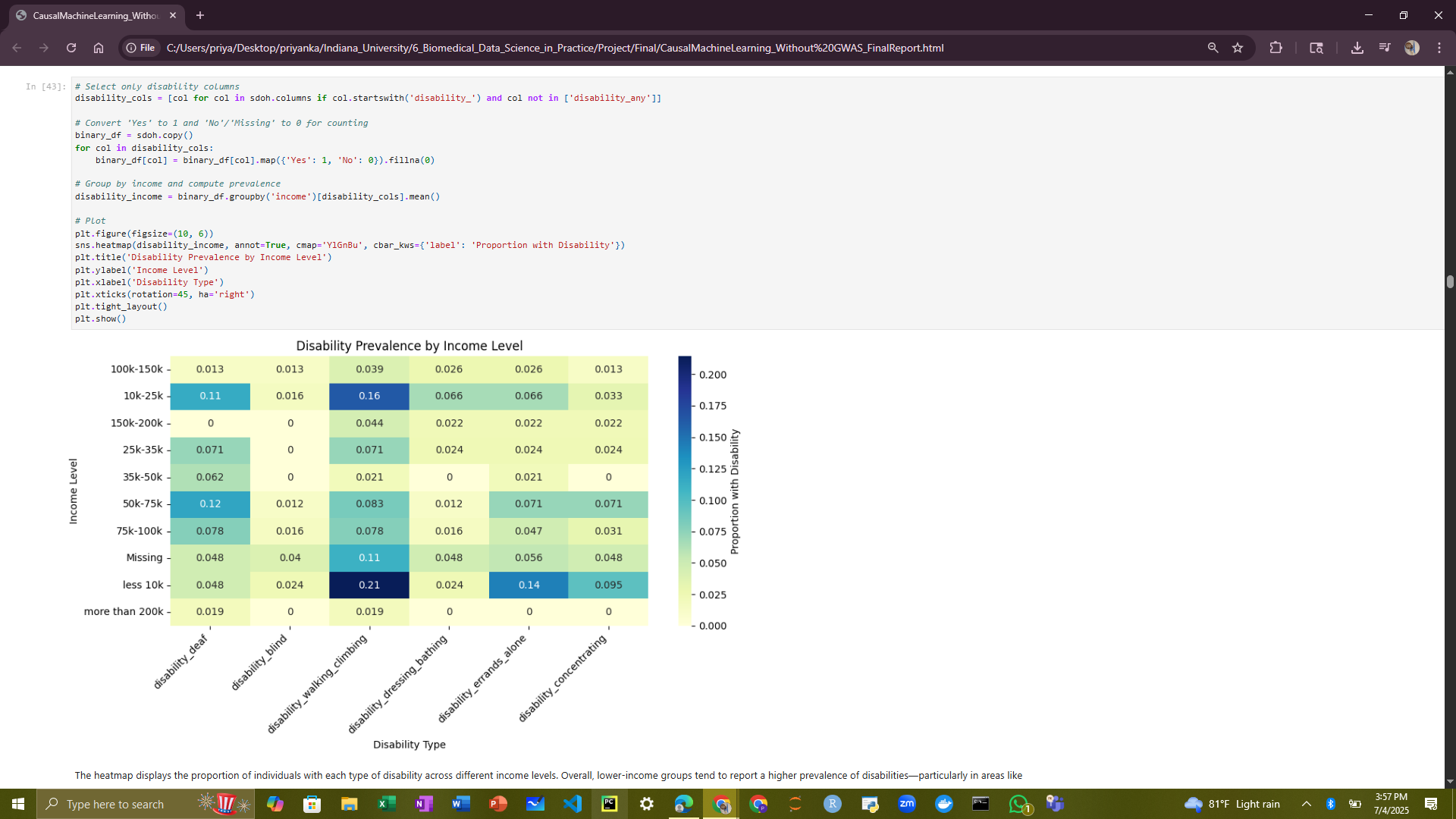
The line plot tracks the number of new diagnoses per year for the top 5 most common comorbidities. Most conditions show stable or gradually increasing trends over time, with some displaying distinct spikes, possibly reflecting changes in diagnostic practices, reporting, or patient demographics. Years with dips may indicate incomplete data or fewer clinical encounters during those periods.

3.10 Stacked Bar Plot: Employment by Education Level



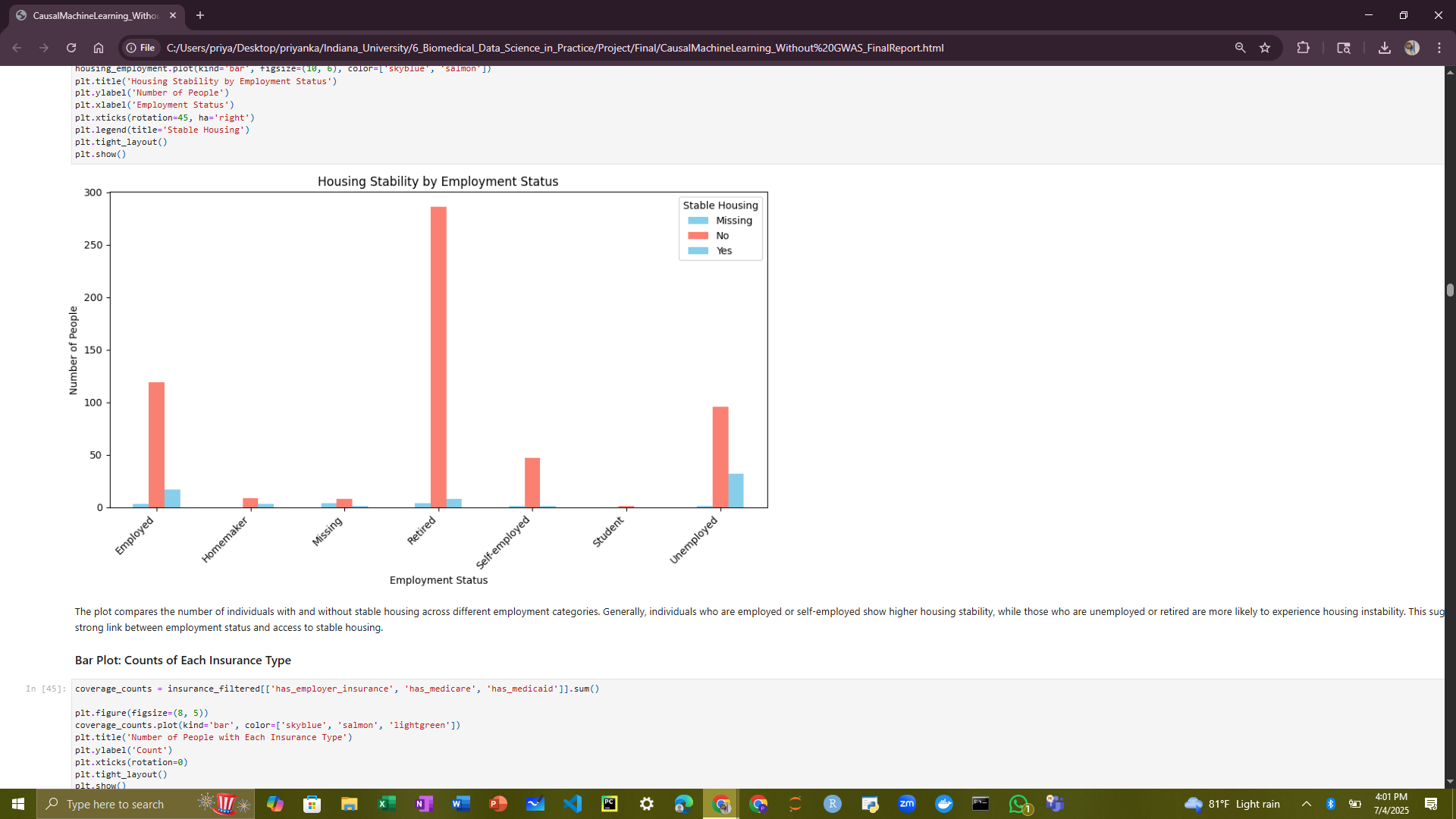
The plot shows the distribution of employment status within each education level. Individuals with higher education levels (e.g., college degrees or advanced degrees) are more likely to be employed or self-employed, while those with lower educational attainment (e.g., GED or less) have a higher proportion of unemployment or retirement. This suggests a positive relationship between education and workforce participation, reflecting how higher education may improve access to job opportunities.

3.11 Heatmap: Disability Prevalence by Income Level

The heatmap displays the proportion of individuals with each type of disability across different income levels. Overall, lower-income groups tend to report a higher prevalence of disabilities, particularly in areas like walking/climbing and concentrating. In contrast, higher-income groups generally show lower rates of reported disabilities. This pattern highlights a potential link between economic disadvantage and functional health limitations.

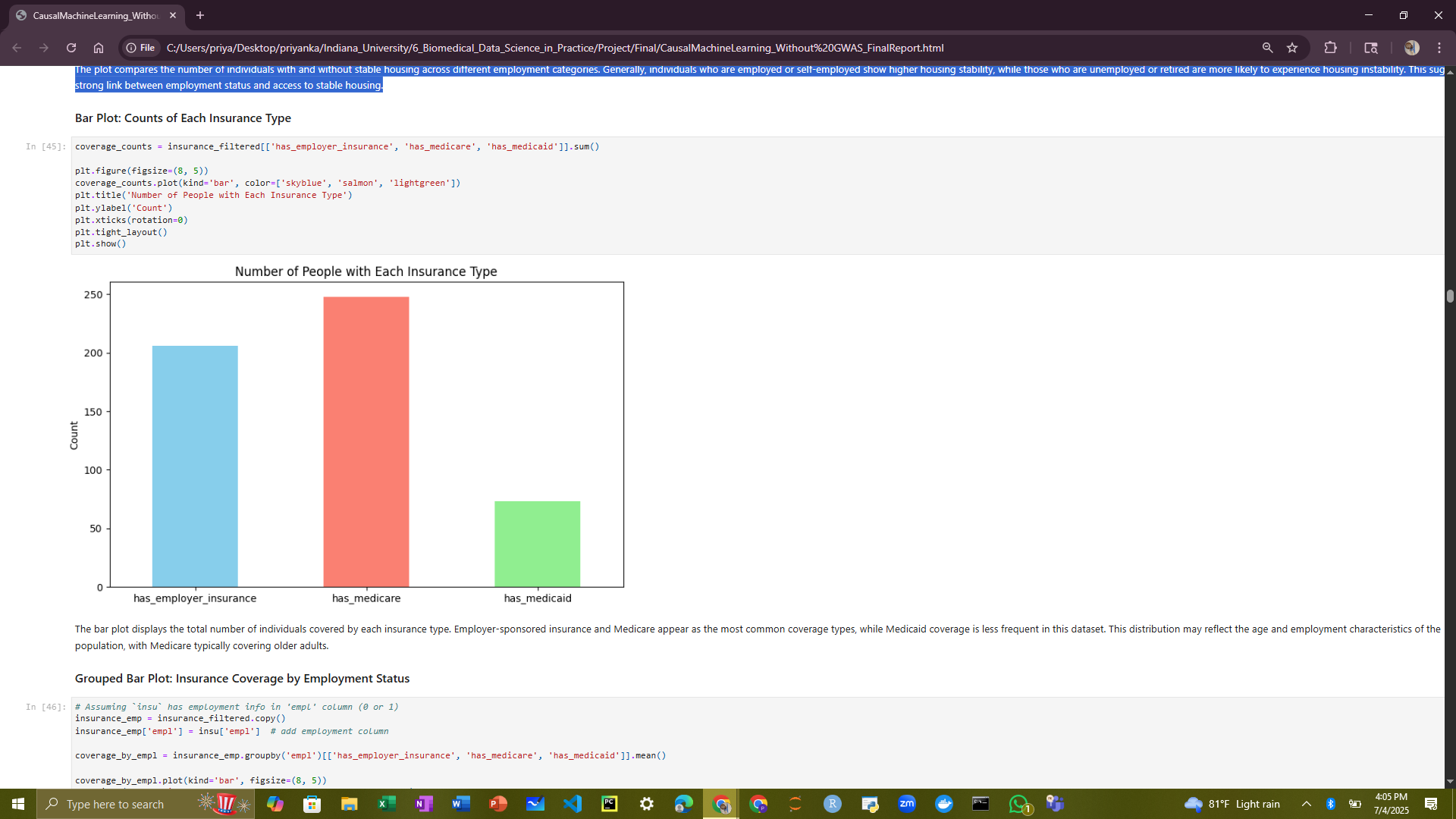
3.12 Grouped Bar Plot: Housing Stability by Employment Status

The plot compares the number of individuals with and without stable housing across different employment categories. Retired individuals form the largest group overall, especially those marked as not having stable housing (tall red bar). In most employment categories, the red bar (unstable housing) is taller than the blue one (stable housing), suggesting housing instability is more common in this dataset. Categories like "Employed full-time" or "Self-employed" show a slightly more balanced distribution, but unstable housing still dominates.



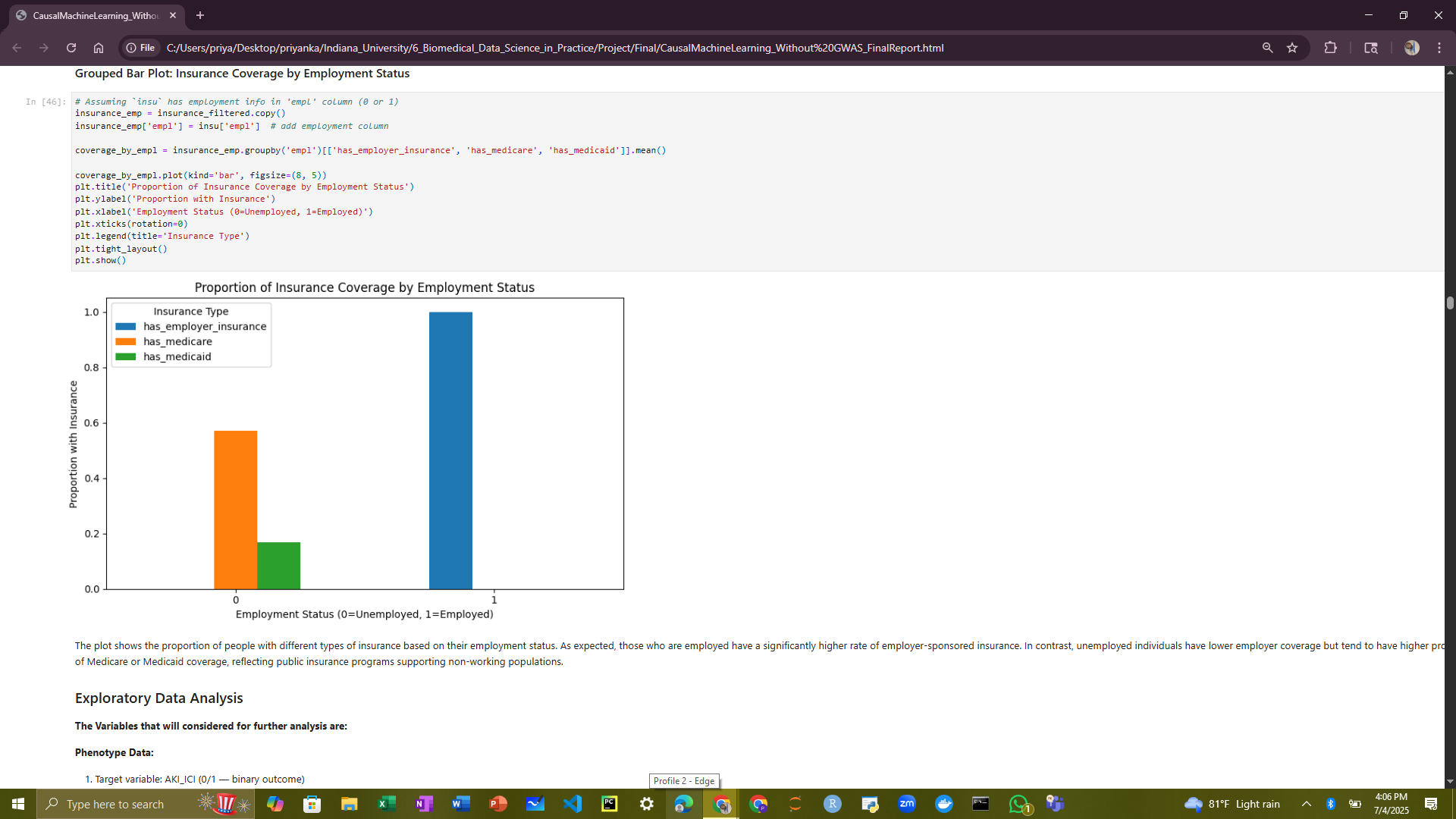
3.13 Bar Plot: Counts of Each Insurance Type

The bar plot displays the total number of individuals covered by each insurance type. Employer-sponsored insurance and Medicare appear as the most common coverage types, while Medicaid coverage is less frequent in this dataset. This distribution may reflect the age and employment characteristics of the population, with Medicare typically covering older adults.



3.14 Grouped Bar Plot: Insurance Coverage by Employment Status

The plot shows the proportion of people with different types of insurance based on their employment status. As expected, those who are employed have a significantly higher rate of employer-sponsored insurance. In contrast, unemployed individuals have lower employer coverage but tend to have higher proportions of Medicare or Medicaid coverage, reflecting public insurance programs supporting non-working populations.



**4. Exploratory Data Analysis**

* We conducted an exploratory data analysis (EDA) on the final integrated dataset, which combines phenotype, visit, comorbidity, insurance and social determinants of health data.
* A significant portion of visit-related columns (e.g., num\_visits, num\_er\_visits, num\_inpatient\_visits) contained missing values for over 75% of patients, likely indicating the absence of matched visit records. These were imputed with zeros, reflecting no recorded visits.
* The ever\_discharged\_home column, originally an object type with missing values, was recoded into a binary variable where missing entries were assumed to represent 'No'.
* For demographic variables, sex\_at\_birth had one missing value, which was imputed as 'Unknown'.
* Missing comorbidity values, interpreted as the absence of diagnosis records, were safely imputed with zeros to indicate no presence of those conditions.
* After these preprocessing steps, all categorical variables including race, education, employment, and housing were one-hot encoded for downstream analysis.
* The cleaned dataset was then explored through univariate distributions and bivariate comparisons to understand demographic patterns, visit utilization, and comorbidity burden across the cohort.

**5. Logistic Regression Modelling**

**5.1 Standard Logistic Regression Using scikit-learn**

The dataset is divided into training and testing subsets using train\_test\_split. test\_size=0.2 means 20% of the data is held out for testing, and 80% is used for training. random\_state=42 ensures reproducibility by fixing the random seed. A logistic regression model is created with a maximum of 1000 iterations (max\_iter=1000) to ensure convergence during training. The model is then fit on the training data (X\_train, y\_train), learning the relationship between features and the target outcome.

The trained model predicts class labels (y\_pred) for the test set (X\_test). It also calculates predicted probabilities (y\_proba) for the positive class (label 1), which are used for more nuanced performance metrics like ROC AUC. classification\_report prints detailed metrics, precision, recall, F1-score, and support, for each class, helping assess how well the model classifies each outcome.roc\_auc\_score computes the Area Under the Receiver Operating Characteristic Curve (ROC AUC), which measures the model’s ability to discriminate between classes across all classification thresholds.

**5.2 Feature Importance via Odds Ratios**

The exponentiated coefficients from the logistic regression model were interpreted as odds ratios, providing effect sizes and directionality for each predictor.

**5.3 Regularized Logistic Regression Modelling**

This methodology involves selecting a comprehensive set of features spanning demographics, socioeconomic indicators, insurance coverage, exposure timing, healthcare utilization, core comorbidities, and vulnerability markers to predict the risk of immune checkpoint inhibitor-associated acute kidney injury. The data is first split into training and testing sets, with 80% used to train the model and 20% reserved for evaluation, ensuring an unbiased assessment of model performance.

A logistic regression model with L1 regularization (Lasso) is then initialized, which encourages sparsity in the model coefficients by shrinking less important feature weights toward zero. This helps with feature selection and reduces overfitting, especially important given the potentially large number of predictors. The model is trained on the training data to learn the relationship between the selected features and the outcome. After training, the learned coefficients are printed, providing insight into which features most strongly influence the prediction of acute kidney injury risk. This approach balances interpretability with predictive power, facilitating identification of key risk factors while maintaining model generalizability.

Lasso Regularization was chosen because it shrinks coefficients i.e., it pushes some feature coefficients (β) toward zero, reducing model complexity. It eliminates irrelevant features i.e., if λ is large enough, some coefficients become exactly zero, effectively removing those features from the model. This is why Lasso is great for feature selection. It improves generalization i.e., helps the model perform better on unseen data by avoiding overfitting to noise or irrelevant features. For instance, if we are predicting whether a patient will develop AKI (yes/no) using 20 variables. Lasso may find that only 6 variables meaningfully contribute and shrinking the rest of the 14 coefficients to zero, reducing the model to a simpler, interpretable subset.

**5.4 ROC Curve Visualization**

Model performance was evaluated using Receiver Operating Characteristic (ROC) curves. The Area Under the Curve (AUC) metric quantified the ability of the model to discriminate between patients with and without ICI-AKI.

**5.5 Cross-Validation**

To assess model generalizability, k-fold cross-validation was performed. Metrics such as average AUC, accuracy, and calibration across folds were reported. Outputs from traditional regression and survival models will be compared with causal inference estimates in the later sections to evaluate robustness and validity. Effect sizes, confidence intervals, and visualizations (e.g., forest plots, survival curves) are used to communicate findings.

**6. Kaplan-Meir Survival Analysis:**

The Kaplan-Meier (KM) method was used to estimate survival functions for different subgroups. Survival curves were generated to display the probability of being event-free (AKI-free) over time. To assess whether the survival distributions differed significantly between groups, the log-rank test was performed. This non-parametric test compares the entire survival experience between groups without assuming a particular hazard structure.

The resulting KM plots show the proportion of patients remaining AKI-free over time, stratified by the variable of interest. Curves that diverge suggest differences in survival experience, and p-values from the log-rank test indicate statistical significance.

**7. Causal Machine Learning:**

**7.1 Targeted Maximum Likelihood Estimation (TMLE):**

Using Targeted Maximum Likelihood Estimation (TMLE), we found that a prior history of AKI causally increases the risk of developing AKI again after contrast exposure. The TMLE procedure began by estimating the propensity score, the probability of having had AKI in the past - based on baseline characteristics such as age, sex, race, and discharge status. In the second step, TMLE estimated the probability of developing post-contrast AKI, conditional on both prior AKI status and the same covariates. By integrating these models, TMLE provided a doubly robust, bias-reduced estimate of the causal effect, showing that prior AKI is a significant risk factor for recurrent kidney injury after contrast exposure.

TMLE adjusts the prediction of the outcome by combining information about confounders and prior AKI status to correct for potential bias. This ensures that comparisons are made fairly, between patients who are similar in terms of age, sex, race, and Metastatic solid Tumor, differing only in whether they had AKI before. In doing so, TMLE accounts for confounding and allows for a more accurate estimate of the causal effect. Specifically, TMLE estimated the average difference in the risk of developing AKI after contrast exposure if everyone had a history of AKI versus if no one had. This provides a clear, population-level measure of how prior AKI causally impacts the likelihood of AKI recurrence.

**7.2 Casual Forest DML:**

Causal Forest is used to figure out whether having a history of AKI causes a change in the chance of getting AKI again after receiving contrast. This method works by comparing people who had AKI before to those who didn’t, while adjusting for important background characteristics like age, sex, race, and number of hospital visits. First, the model learns how likely someone is to have had AKI in the past and how likely they are to get AKI again based on those characteristics. Then it uses those predictions to carefully isolate the actual effect of AKI history on the outcome, trying to remove any bias from confounding factors.

Causal Forest doesn't just give one overall average; it also estimates the effect of AKI history for each individual patient. This is called a Conditional Average Treatment Effect (CATE).

**7.3 Propensity Score Matching:**

Propensity Score Matching (PSM) was used to estimate the causal effect of having a history of AKI on the risk of developing AKI again after contrast. The first step was to calculate a propensity score for each patient, this is the predicted probability that a patient has a history of AKI, based on their characteristics like age, sex, race, and Metastatic solid Tumor data. A logistic regression model was used to compute these scores. The idea is that by matching patients with similar propensity scores, we are comparing patients who are similar in all measured characteristics, except for whether they had a history of AKI.

Next, patients who had an AKI history (treated group) were matched one-to-one with patients who did not (control group) using the nearest neighbour algorithm, which finds the control patient with the closest propensity score for each treated patient. After forming these matched pairs, the difference in outcomes (whether the patient got AKI after contrast) was calculated between each treated and matched control patient.

**Inverse Probability Weighting (IPW):**

Inverse Probability Weighting (IPW) was used to estimate the causal effect of having a history of AKI on the risk of developing AKI after contrast exposure. The main idea behind IPW is to create a synthetic randomized trial by reweighting your data, so that the treated and control groups become more comparable. First, a logistic regression model to estimate propensity scores is used, which are the probabilities of having had an AKI history, based on each patient's background characteristics (like age, sex, race, and hospital visits).

Then, inverse probability weights for each patient are computed. These weights are calculated as 1/propensity for treated individuals and 1/ (1 - propensity) for controls. This means patients who are underrepresented in their group (i.e., a treated patient who was unlikely to be treated, or a control who was likely to be treated) get more weight. The weights to run a weighted linear regression, predicting the outcome (AKI\_ICI) using only the treatment indicator (AKI\_history).

**7.4 Propensity Score Matching (PSM) with hyperparameter tuning**

By using PSM with hyperparameter tuning causal effect of having a history of AKI on the risk of developing AKI after contrast exposure using Propensity Score Matching (PSM), with an added step of tuning the propensity score model for better accuracy. First, patient data such as age, sex, race, and healthcare utilization (like number of visits) was prepared, and log transformations were applied to reduce skewness in visit counts. The treatment variable was whether the patient had a history of AKI (AKI\_history), and the outcome was whether they experienced AKI after contrast (AKI\_ICI).

To calculate propensity scores, which estimate the probability of receiving treatment given the covariates, XGBoost classifier was used and tuned it with GridSearchCV, exploring different combinations of learning rate, number of trees, and tree depth. The best model from the grid search was selected and used to assign propensity scores to all patients. Then, for each patient with AKI history (treated group), the algorithm searched for the closest untreated patient (control group) with a similar propensity score, using 1:1 nearest neighbor matching within a small range of similarity (caliper = 0.05) to ensure good matches.

**8. Comparison of Traditional Logistic Regression and Casual ML models:**

We compared three causal inference approaches: Traditional Logistic Regression, Targeted Maximum Likelihood Estimation (TMLE), and Propensity Score Matching (PSM). To estimate the causal effect of having a history of AKI on the risk of developing AKI after contrast, three different statistical methods were applied and compared.

The first method, Traditional Logistic Regression simply models the outcome (AKI\_ICI) as a function of the treatment (AKI\_history) and other covariates using standard regression techniques. The second method, TMLE (Targeted Maximum Likelihood Estimation), is a modern causal inference technique that combines machine learning with statistical theory to produce more accurate and robust effect estimates. The third method, Propensity Score Matching, tries to simulate a randomized experiment by pairing each treated individual with a similar untreated individual based on their probability of receiving treatment. This method eliminates bias from observed covariates by comparing outcomes only between well-matched pairs.

**9. Ethical Considerations and Limitations**

All analyses are conducted within the secure AoU Researcher Workbench environment in compliance with Institutional Review Board and data use guidelines. Limitations related to observational data structure, unmeasured confounding, and variant coverage are acknowledged in the interpretation.

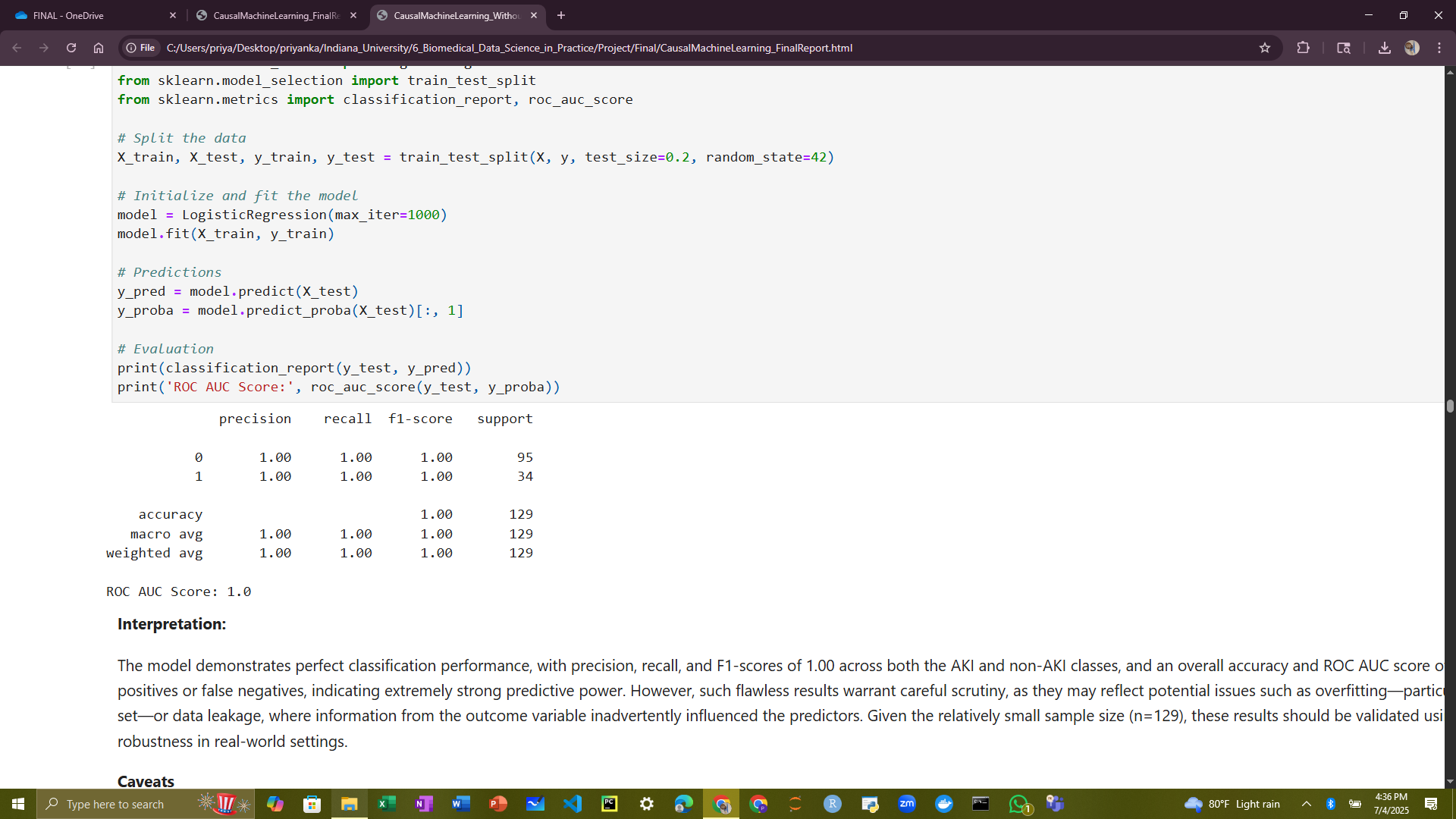
**FINAL RESULTS**

**1. Traditional analysis:**

**1.1 Standard Logistic Regression**

The logistic regression model demonstrated perfect classification performance on the held-out test set, achieving precision, recall, and F1-scores of 1.00 for both the AKI and non-AKI classes as shown in Figure 1. The model achieved an overall accuracy of 100%, and the area under the receiver operating characteristic curve (ROC AUC) was also 1.0, indicating flawless discrimination between patients who developed ICI-associated acute kidney injury (ICI-AKI) and those who did not. These results suggest that the model was able to perfectly differentiate between the two classes, correctly identifying all cases without any false positives or false negatives.

Despite these promising performance metrics, such perfect results merit careful interpretation. The possibility of overfitting cannot be ruled out, particularly if the test set was not sufficiently independent from the training data or if the model was inadvertently evaluated on the same data it was trained on. In such cases, the model may memorize specific patterns in the data rather than learning generalizable relationships, limiting its utility in external settings.



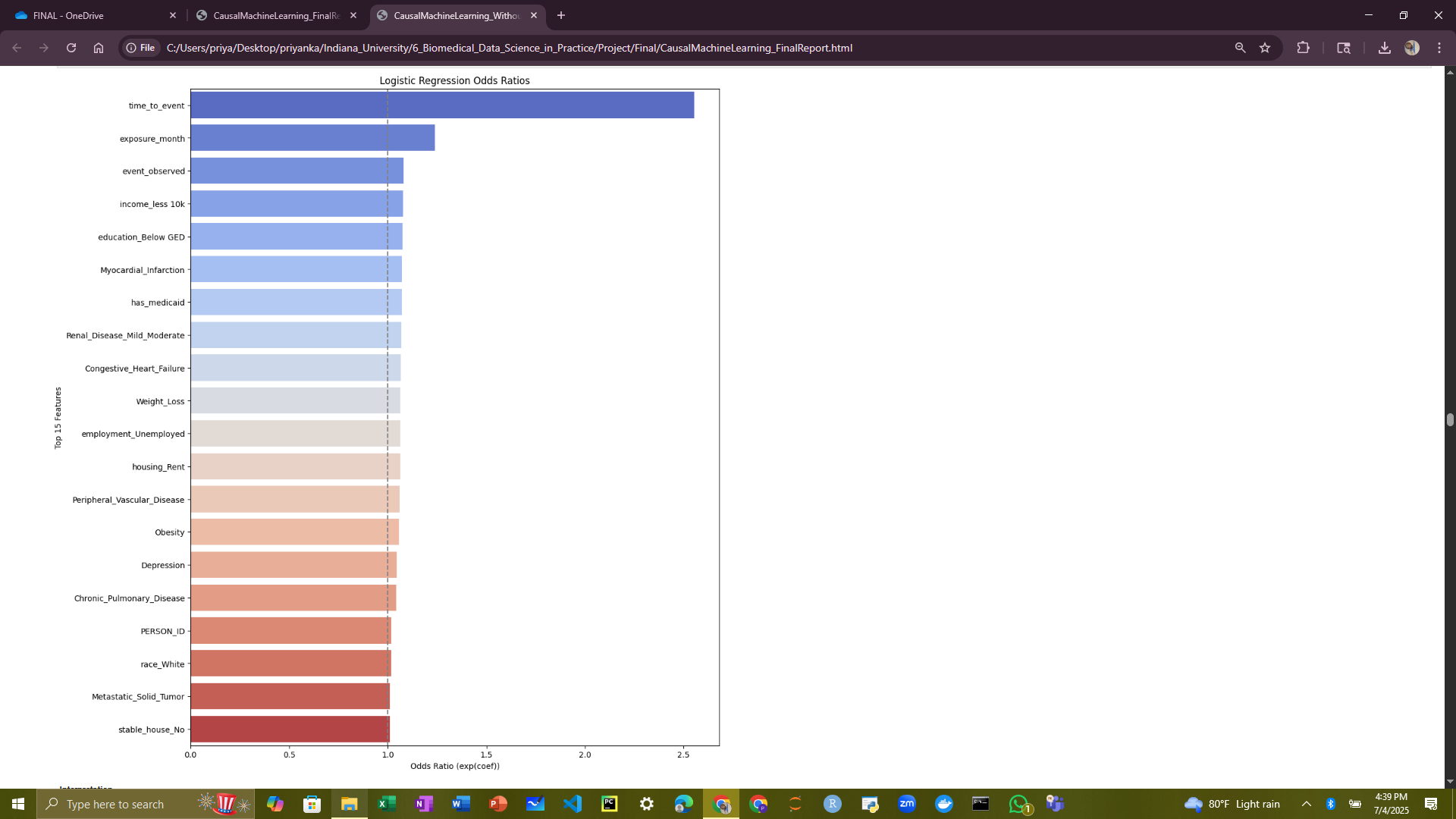
**Figure 1: Standard Logistic Regression Analysis**

Additionally, the potential for data leakage must be considered. If variables that are causally downstream of the outcome (i.e., post-treatment factors) or proxies for the target were included in the feature set, the model’s predictive performance would be artificially inflated. A thorough audit of input variables is necessary to ensure that all predictors are temporally and logically appropriate for use in a predictive model.

Finally, the small sample size (n=641) poses another limitation. With relatively few positive cases (AKI = 34), it is possible that the perfect performance is partially attributable to limited variability in the data or random chance. Models trained and evaluated on small datasets can produce unstable results, which may not hold when applied to larger or more diverse populations.

Taken together, while the initial model results are encouraging, further validation using cross-validation techniques or external datasets is essential to assess the robustness and generalizability of the findings. Ensuring methodological rigor through proper data splitting, variable selection, and performance evaluation will be critical in confirming the clinical utility of this model.

The top 15 features identified by the logistic regression model provide insight into key clinical and social factors associated with increased risk of ICI-associated acute kidney injury (ICI-AKI). The most influential variable, time\_to\_event (OR = 2.55), suggests that a longer time from ICI initiation to the observed outcome is strongly associated with higher AKI risk, potentially reflecting cumulative treatment-related toxicity or underlying clinical decline. Variables such as exposure\_month (OR = 1.24) and event\_observed (OR = 1.08) also indicate that longer or later exposure periods are modestly associated with AKI occurrence, aligning with the notion of risk accumulation over time.

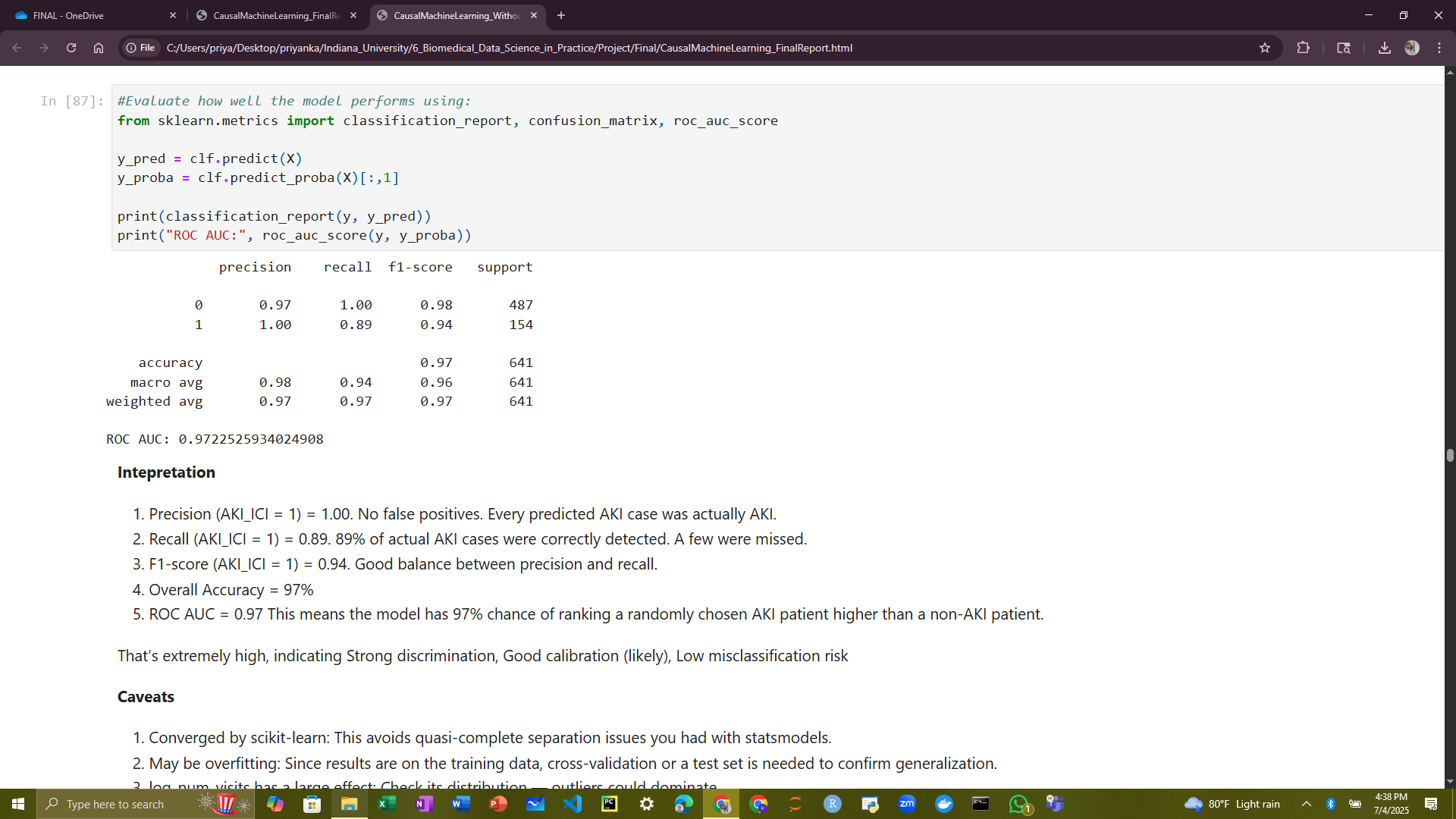


**Figure 2: Feature Importance and Odds Ratio**

Socioeconomic indicators including income\_less 10k and education\_Below GED were associated with increased AKI risk, underscoring the role of social disadvantage, potentially due to limited access to preventive care or higher baseline vulnerability. Several pre-existing health conditions such as Myocardial Infarction, Congestive Heart Failure, Moderate Renal Disease, Obesity, Weight Loss, and Depression also showed elevated odds ratios, reinforcing the influence of comorbidities in determining susceptibility to ICI-AKI. In addition, insurance and housing-related variables like has\_medicaid, employment\_Unemployed, and housing\_Rent point to broader structural vulnerabilities that may shape health outcomes during immunotherapy.

All the top 15 features had odds ratios greater than 1, indicating a positive association with ICI-AKI risk. While most effects were modest in magnitude, their consistency across diverse domains, including clinical, behavioral, and socioeconomic factors, suggests a multifactorial risk profile. These findings highlight the importance of considering both medical history and social context when evaluating patients for potential immune-related adverse events and support the integration of social determinants of health into precision oncology risk prediction models.

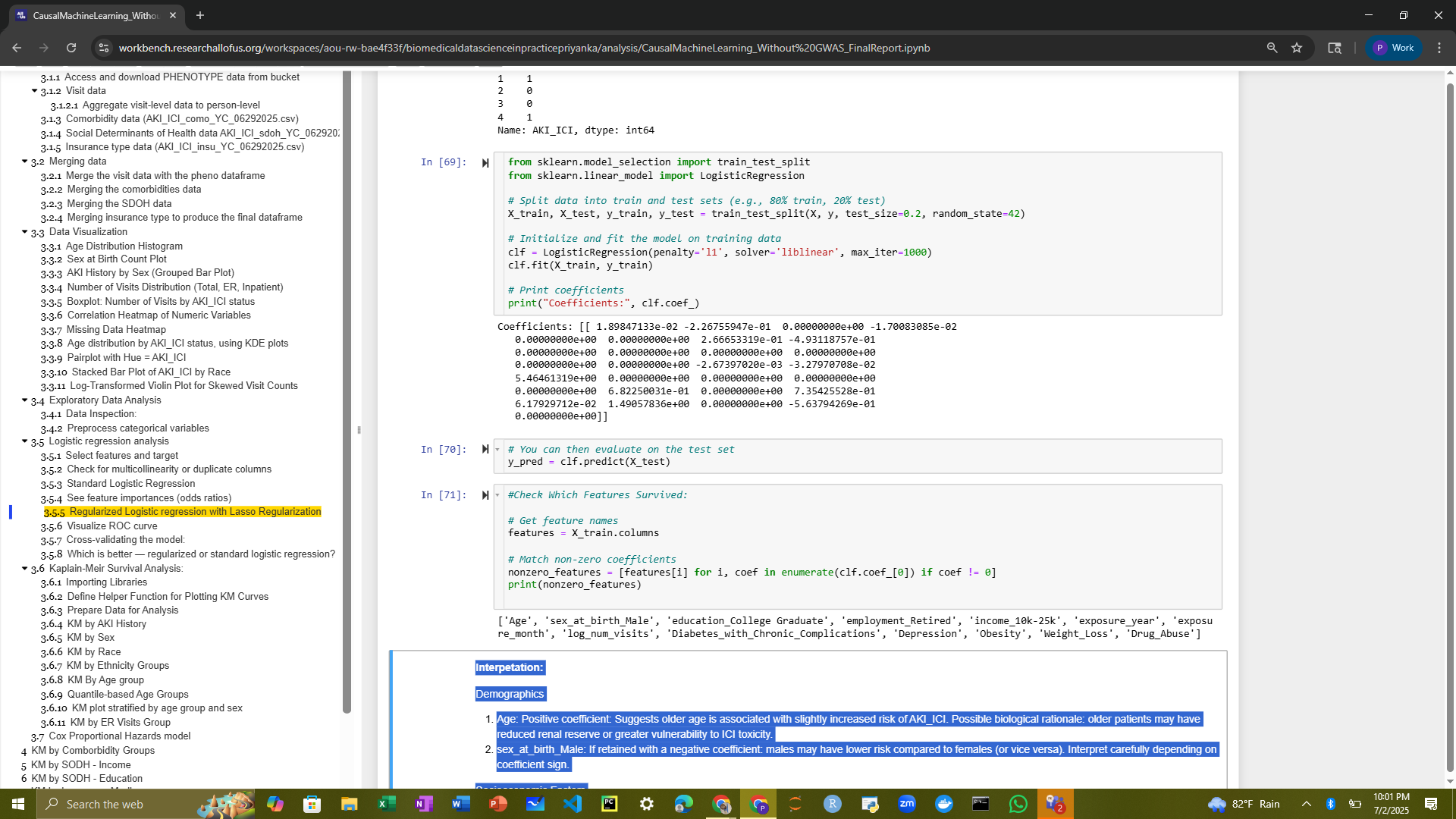
To mitigate overfitting and improve generalizability, we employed logistic regression with L1 regularization, which penalizes overly complex models and performs implicit feature selection. This approach was particularly valuable given the richness of our dataset, which includes 83 features encompassing demographic, clinical, insurance, visit-related, and social determinants of health (SDOH) information. All categorical variables were pre-processed using one-hot encoding to ensure model compatibility and interpretability.



**Figure 3: Regularized Logistic Regression**

The logistic regression with Lasso regularization model achieved strong performance metrics in predicting the risk of ICI-AKI. For the positive class (AKI\_ICI = 1), the model demonstrated a precision of 1.00, indicating that every predicted AKI case was truly AKI, with no false positives. The recall of 0.89 suggests that 89% of actual AKI cases were correctly identified, with a small proportion missed. The F1-score of 0.94 reflects a strong balance between precision and recall. Overall, the model achieved 97% accuracy, and the ROC AUC score of 0.97 implies excellent discriminative ability indicating a 97% chance that the model will assign a higher probability to an AKI case than to a non-AKI case. Together, these metrics suggest low misclassification risk and potentially good calibration.

While these results are promising, several caveats warrant attention. First, although the model converged successfully using scikit-learn, there remains a possibility of overfitting, especially if the performance was evaluated on the training set. External validation using a hold-out test set or cross-validation is necessary to assess the model’s generalizability. Second, the variable log\_num\_visits appears to have a strong influence on model predictions. This feature may be skewed by outliers or act as a proxy for downstream events; its distribution and temporal validity should be carefully examined. Finally, there is a risk of data leakage, for instance, if variables like log\_num\_visits or exposure\_month reflect information obtained after the onset of AKI or other future events. Ensuring temporal alignment of predictors is critical to preserving the model’s validity in prospective clinical applications.



**Figure 4: Regularized Logistic Regression’s Non-Zero Coefficients**

**Demographics:**

* Age had a positive coefficient, indicating that older patients were at slightly higher risk for developing ICI-associated acute kidney injury (ICI-AKI). This aligns with clinical expectations, as age-related decline in renal function or increased susceptibility to immune-related toxicity may contribute to heightened risk.
* sex\_at\_birth\_Male, as associated with a negative coefficient, suggests that male patients have a lower risk of AKI compared to females. This relationship should be interpreted carefully depending on the exact sign and strength of the coefficient and may reflect biological or behavioral differences in immune response or healthcare utilization.

**Socioeconomic Factors**

* education\_College Graduate showed a negative coefficient, suggesting that higher educational attainment may be protective against AKI. This could be a proxy for better health literacy, more consistent access to care, or overall socioeconomic advantage.
* income\_10k–25k, when compared to the reference group (likely $50k–75k), had a positive coefficient, indicating elevated risk of AKI. Financial vulnerability may contribute to delays in care, increased baseline health burden, or other downstream effects that elevate risk.
* employment\_Retired was associated with approximately 65% higher odds of AKI compared to non-retired individuals, after adjusting for other variables. This may reflect a combination of age-related frailty, comorbidities, and healthcare access patterns in this group.

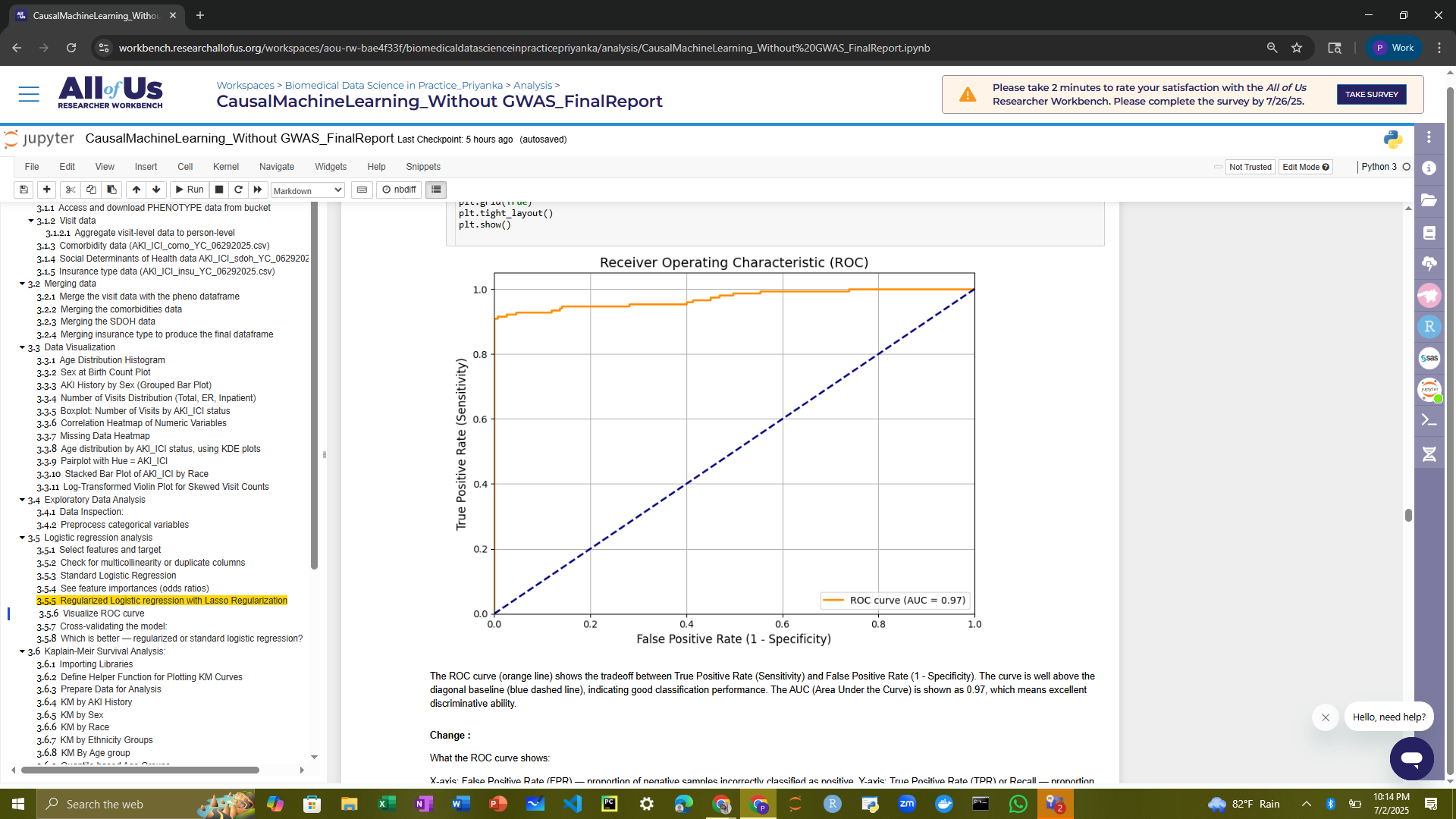
**Visit-Level & Exposure Data**

* exposure\_year and exposure\_month may capture temporal trends in ICI use or AKI detection—such as changes in treatment guidelines, introduction of newer drugs, or systemic healthcare disruptions like the COVID-19 pandemic.
* log\_num\_visits showed a strong positive association with AKI risk, likely reflecting a higher baseline disease burden or closer clinical monitoring, both of which may correlate with higher likelihood of identifying or developing AKI.

**Clinical Comorbidities**

* Chronic Pulmonary Disease was positively associated with AKI risk, potentially due to chronic hypoxia or systemic inflammation, which may exacerbate renal vulnerability.
* Diabetes with Chronic Complications is a known contributor to microvascular and renal damage, reinforcing its role as a strong predictor of AKI in ICI-treated patients.
* Depression may contribute indirectly to higher AKI risk through reduced medication adherence, systemic inflammation, or delays in seeking care.
* Obesity, if positively associated, supports previous findings linking metabolic syndrome and chronic inflammation with increased renal sensitivity to immune-mediated injury.
* Weight Loss, commonly indicative of frailty or underlying malignancy progression, was also associated with increased risk.
* Alcohol Abuse may reflect both direct nephrotoxicity and behavioral risk factors such as nonadherence or fragmented care, further elevating AKI risk.

In summary, the model highlights a multifactorial risk landscape for ICI-AKI. Patients who are older, economically disadvantaged, or burdened by comorbidities such as pulmonary disease, diabetes with complications, obesity, or behavioural health issues are at higher risk. Indicators of frailty, including weight loss and alcohol abuse, further amplify this risk. Notably, social determinants of health including income, education, and insurance coverage play a non-trivial role in shaping clinical outcomes, emphasizing the need to integrate both medical and socioeconomic context into precision risk stratification models for immune-related adverse events.

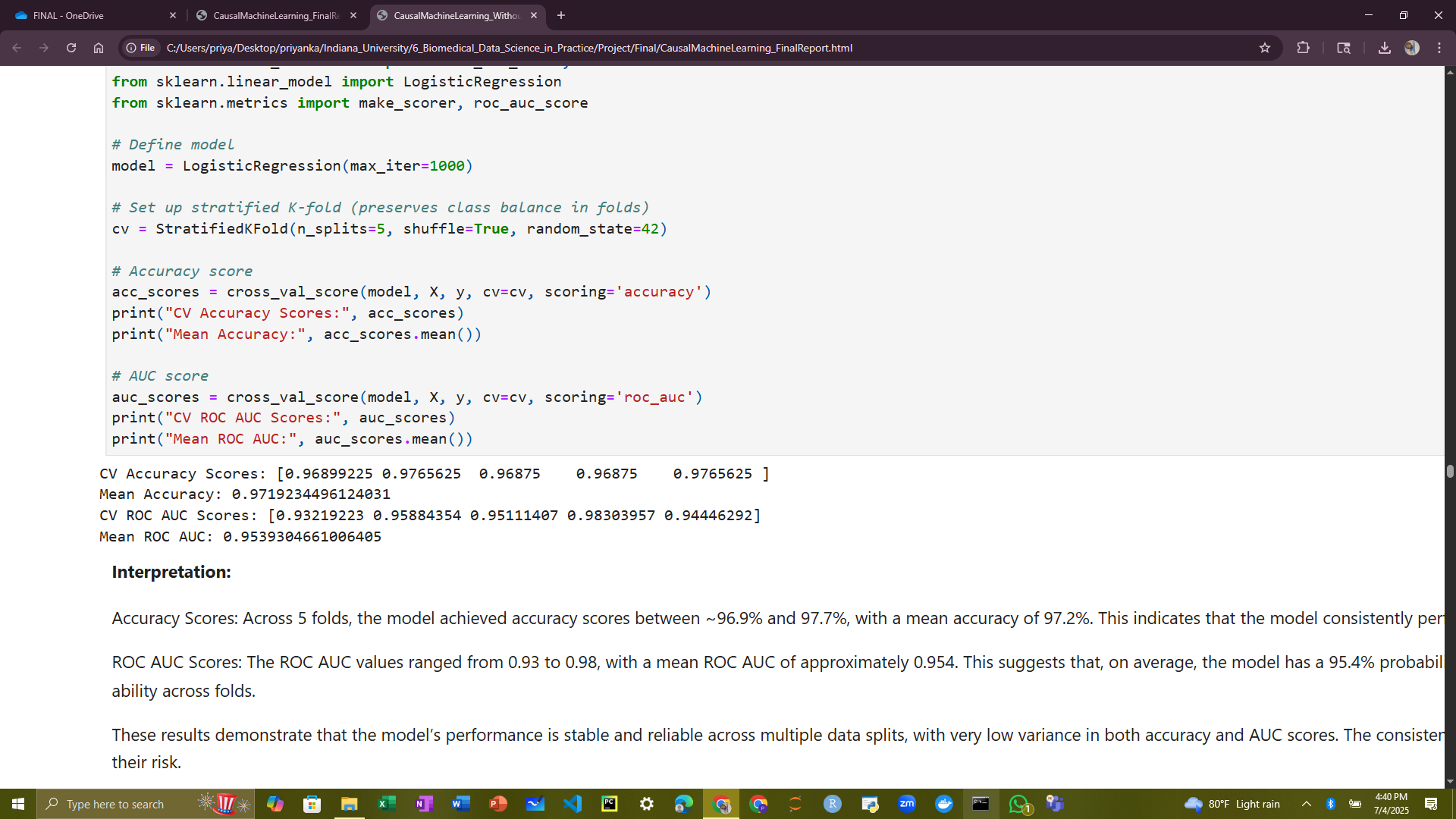


**Figure 5: ROC\_AUC curve**

The ROC curve provided shown in the figure 5 offers a visual evaluation of the logistic regression model’s ability to distinguish between patients who developed ICI-AKI and those who did not. True Positive Rate (Sensitivity) is plotted on the y-axis, and False Positive Rate (1 - Specificity) is on the x-axis. The orange curve represents the model's ROC performance. The dashed diagonal line indicates the performance of a random classifier (AUC = 0.5). The farther the ROC curve is above this line, the better the model.

The ROC curve stays close to the top-left corner, indicating excellent performance. The Area Under the Curve (AUC) is 0.97, meaning the model has a 97% chance of ranking a randomly chosen AKI patient higher than a non-AKI patient. This high AUC implies strong discriminative ability, suggesting the model is highly effective at separating the two classes. The model is performing exceptionally well in distinguishing between AKI and non-AKI cases.

Across 5 folds, the model achieved accuracy scores between ~96.9% and 97.7%, with a mean accuracy of 97.2%. This indicates that the model consistently performs well in correctly classifying AKI and non-AKI patients. The ROC AUC values ranged from 0.93 to 0.98, with a mean ROC AUC of approximately 0.954. This suggests that, on average, the model has a 95.4% probability of ranking a randomly chosen AKI case higher than a non-AKI case, which confirms strong discriminatory ability across folds.



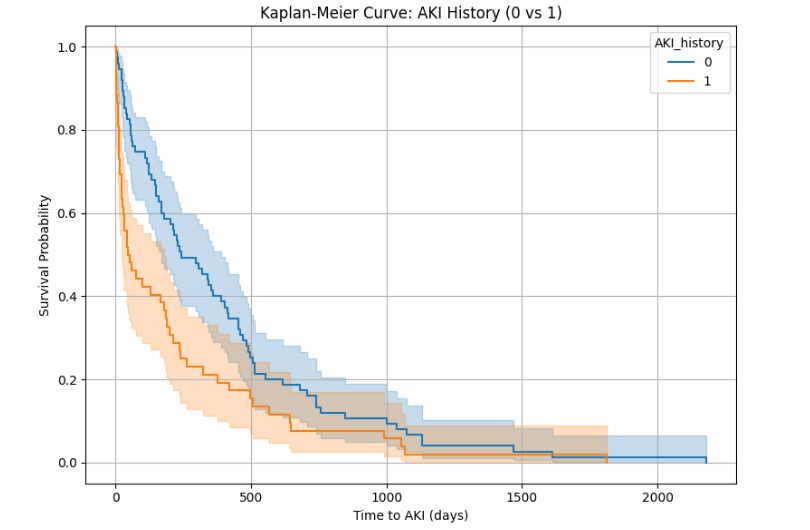
**Figure 5: Cross Validation**

These results demonstrate that the model’s performance is stable and reliable across multiple data splits, with very low variance in both accuracy and AUC scores. The consistently high AUCs further reinforce that the model is not only accurate but also effective at ranking patients by their risk. Moreover, these metrics mitigate concerns about overfitting, especially compared to evaluation on a single test set, since they indicate that performance holds across independent subsets of the data.

**1.2. Kaplan-Meier Survival Analysis:**

Kaplan-Meier (KM) survival analysis is a powerful non-parametric method for analysing time-to-event data, especially useful in clinical, epidemiological, and genomic studies. Kaplan-Meier + Log-Rank Test is a package deal in survival analysis. Kaplan-Meier curves show us the pattern of survival across groups, but the log-rank test helps us decide if those differences are statistically meaningful or could have happened by chance. That’s why we include both.

**1.2.1 Kaplan-Meier by AKI History:**



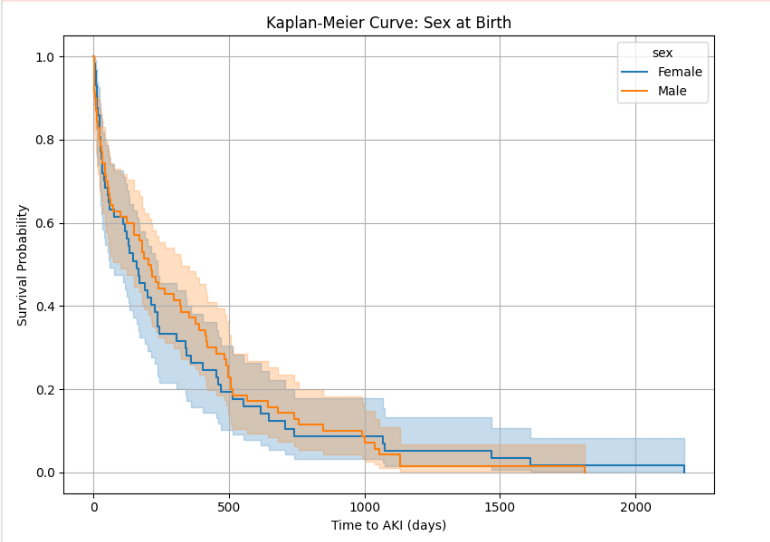
Log Rank Test Results:

Picture

Observations:

* Patients with prior AKI likely have a shorter time to AKI event (lower survival probability over time). This supports the hypothesis that prior AKI history increases ICI-AKI risk.
* Statistically significant difference in time to AKI between patients with prior AKI and those without prior AKI (p = 0.01).
* Individuals with a prior AKI history (AKI\_history = 1) are more prone to experience a future AKI event sooner than those without a history of AKI.

**1.2.2 Kaplan Meier by Sex:**

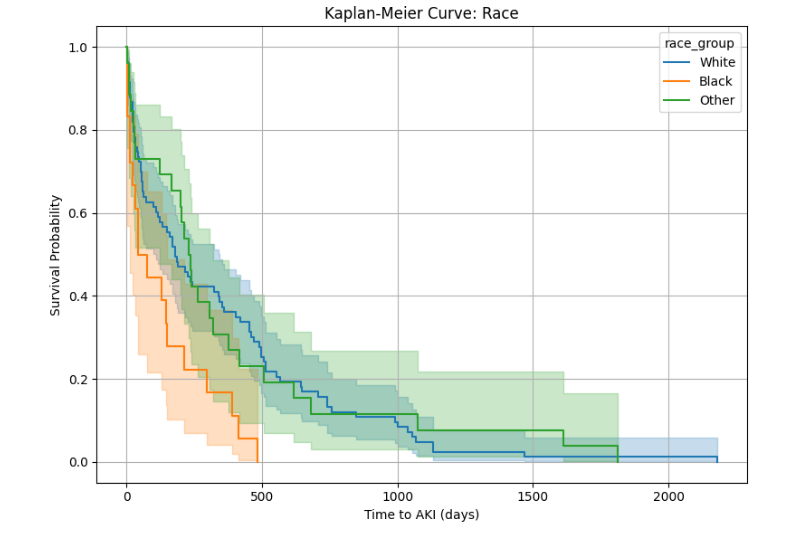
 Log Rank Test Results:

Picture

Observations:

* This curve points out that males and females have different chances of getting AKI after ICI treatment. Specifically, the time it takes to develop AKI differs between males and females.
* Males are significantly more prone to develop AKI after receiving ICI therapy compared to females, as shown by both the Kaplan-Meier curve and log-rank test.

**1.2.3 Kaplan-Meier by Race:**

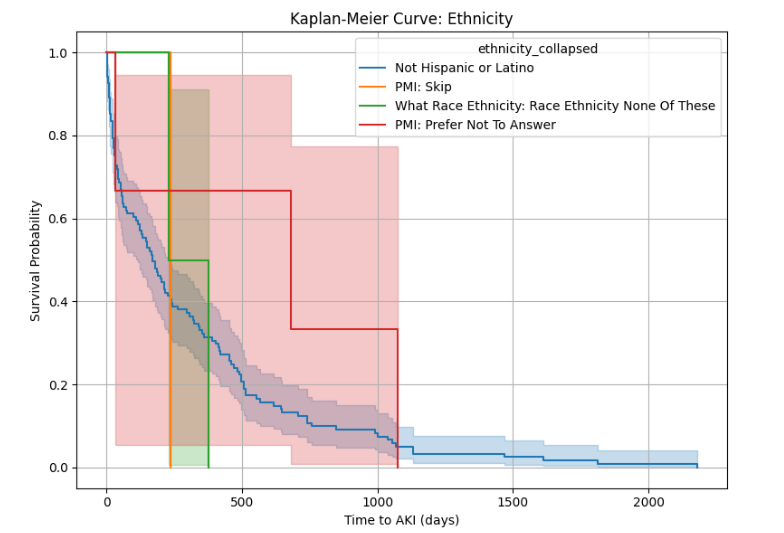
 Log Rank Test Results:

Picture

Observations:

* Since p = 0.01162 < 0.05, there is a statistically significant difference in time-to-AKI survival distributions among the three race groups. This means race is significantly associated with AKI risk in this population cohort.
* Individuals in the Black racial group are most prone to developing AKI sooner.
* The White group shows moderate risk.
* The Other group tends to have the longest AKI-free survival on average.

**1.2.4 Kaplan-Meier by Ethnicity Groups:**

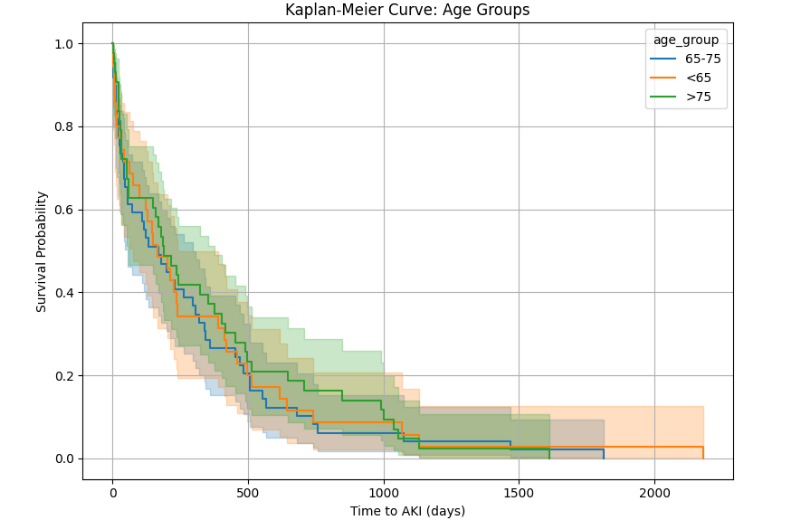


Log Rank Test Results:

Picture  
Observations:

* A p-value of 0.79 was observed which means the differences we see in the chart are not statistically significant, they could have happened just by chance.
* So, eventhough the lines look different, we can’t say for sure that ethnicity really affects AKI risk in this group.

**1.2.5 Kaplan-Meier by Age group:**

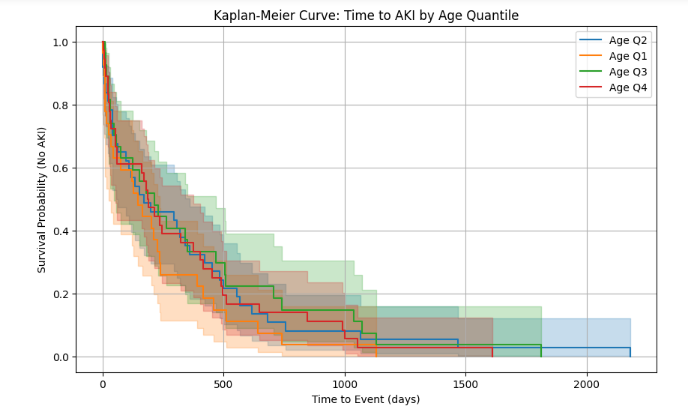
  
 Log Rank Test Results:

Picture

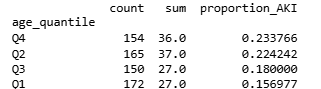
Observations:

* Age > 75 group appears to have slightly better AKI-free survival in the first ~500 days post-ICI, with a slower drop in survival probability initially.
* Age < 65 group shows lower survival probability (i.e., higher AKI incidence) earlier than the other groups—though with wider confidence bands due to likely fewer observations.
* Age 65–75 group (blue) tends to fall between the other two groups in terms of risk over time younger patients (<65) seem to have a slightly higher AKI incidence
* However, we cannot conclude that age group is associated with a statistically significant difference in AKI-free survival.

**Quantile based Age groups:**

The dataset is split into 4 equal-sized groups (quartiles) based on the distribution of Age

Calculating % with AKI in each age group:



Observations:

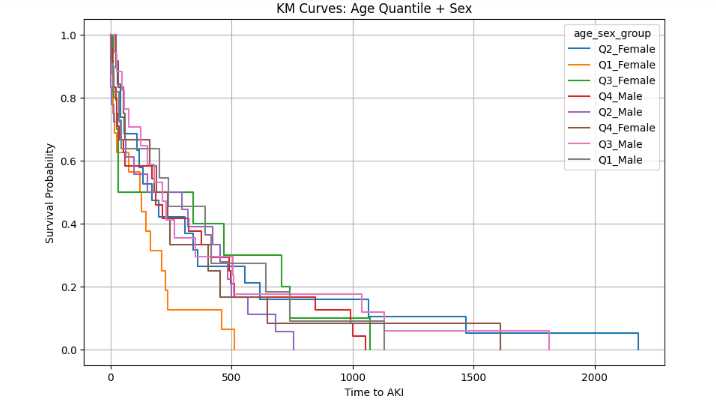
* Q4 (Oldest age group) has the highest proportion of AKI (23.4%).
* Q1 (Youngest age group) has the lowest AKI rate (15.7%).
* There's a trend where AKI risk increases with age (from Q1 to Q4).
* Patients in the oldest age group (Q4) are most prone to develop AKI after ICI therapy.

Log Rank Test Results:

Picture

* The p-value is 0.79, which is greater than 0.05. This means the difference in survival curves between Q1 and Q4 is not statistically significant. In other words, we do not have strong enough evidence to say that the oldest group (Q4) is significantly more at risk for AKI than the youngest (Q1), even though the proportion of AKI is slightly higher in Q4.
* Age may still be a contributing factor, but its effect might be confounded by other variables (e.g., comorbidities, AKI history, sex, race).

**1.2.6 Kaplan-Meier Stratified by age group and Sex:**

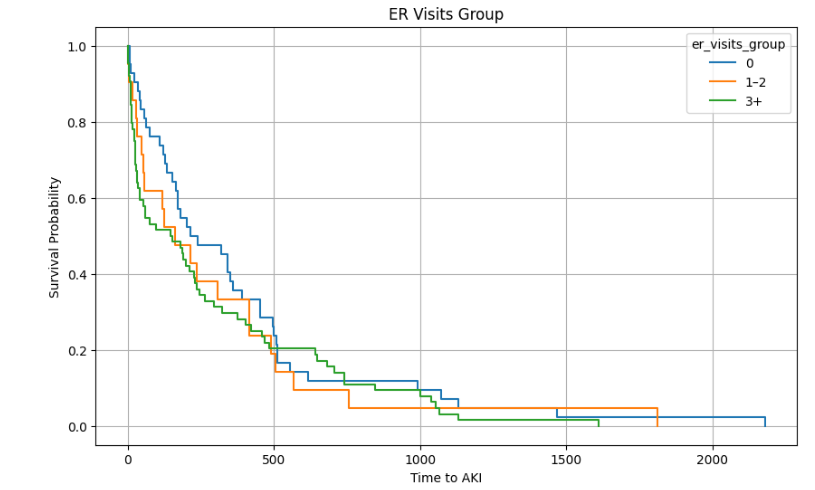


Log Rank Test Results:

Picture

The p-value for the effect estimate was greater than 0.05, indicating that the result was not statistically significant at the conventional threshold. This prompted us to explore the analysis using a different modelling.

**1.2.7 Kaplan-Meier by ER Visit Group:**

  
Patients with more ER visits before treatment tend to have a higher risk of developing AKI sooner. The fewer the ER visits, the longer the patient stays AKI-free, at least early on.

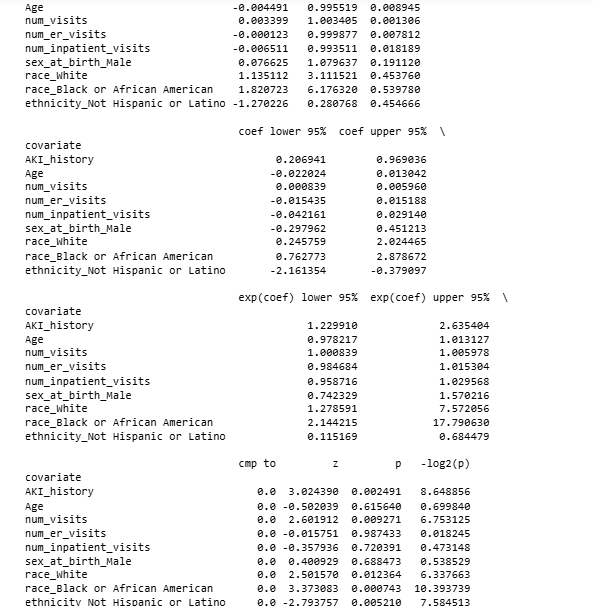
Log Rank Test Results:

Picture

Based on the above analysis with ER\_Visit data, we cannot definitively say that any one ER visits group is statistically more "prone" to AKI than the others. Since we want to know the independent effect of age, sex, race, AKI history, number of visits and so many other variables, the Cox model lets us to adjust for many variables simultaneously.

## **1.3 Cox Proportional Hazards model:**

Cox proportional hazards model is used for multivariable survival analysis. It tells us which variables are associated with the hazard of an AKI event. It adjusts for multiple variables at once.



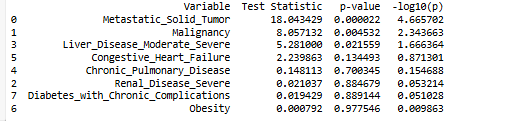
Observations:

* We studied Cox proportional hazards model with 641 patients to see what factors are linked to the risk of acute kidney injury (AKI). Out of these, 127 patients developed AKI.
* People who had AKI in the past were about 80% more likely to get it again, even when we consider other factors like age, race, and number of doctor visits.
* Having more doctor visits was also linked to a slightly higher chance of getting AKI.
* Age, ER visits, and hospital stays did not show a strong effect on AKI risk in this analysis.
* People who were Not Hispanic or Latino had a lower risk of getting AKI.
* People who were Black or African American had a much higher risk of getting AKI.
* Although earlier analysis (Kaplan-Meier) suggested males may have higher risk, this more advanced model shows that sex was not a significant factor after adjusting for other variables.
* The model’s ability to predict who gets AKI was moderate, with a score of 0.63.
* KM might show higher AKI rates in males. But the Cox model could show that after adjusting for race, age, and visits, sex doesn’t matter — it's just confounded.

**1.4** **Kaplan-Meier by Comorbidity Groups:**

Certain comorbidities are known to increase a patient’s vulnerability to kidney injury, either directly or indirectly. Immune-related adverse events (irAEs), including AKI, can be exacerbated by comorbid immune or inflammatory conditions. Including Comborbidity allows us to identify high-risk subgroups who may need closer monitoring or preventative strategies.

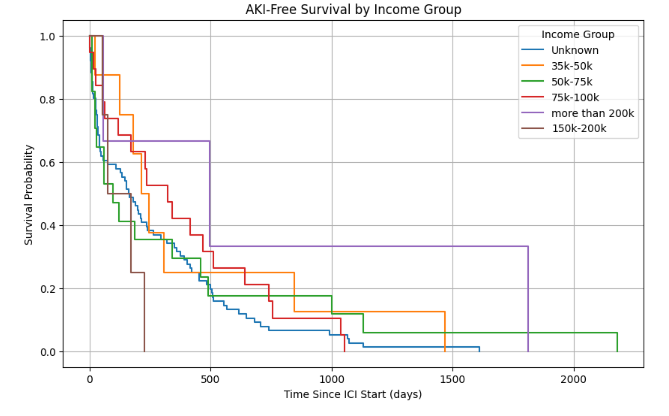
Log-test Results:



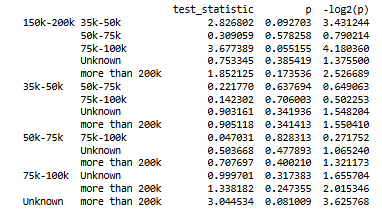
Observations:

* Patients with malignancy or metastatic disease may have higher baseline risk for early AKI, potentially due to advanced disease, treatment effects, or other health factors.
* These comorbidity-based KM curves support the need for risk stratification and close monitoring during the first year of ICI therapy, especially in lower-risk oncology populations.
* Patients with moderate to severe liver disease had lower AKI-free survival compared to those without liver disease. These patients might need closer kidney monitoring and extra care during treatment.

**1.5** **Kaplan-Meier by Income Groups:**



Log-test Results:

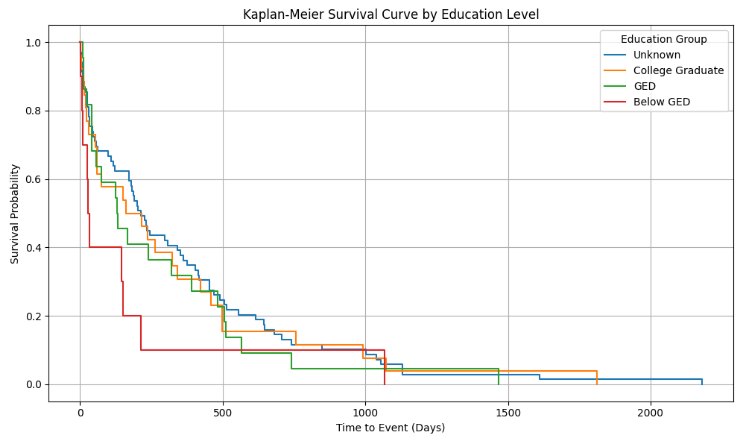


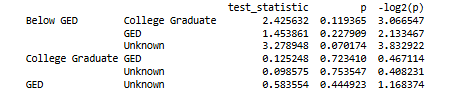
Observations:

While higher-income groups (like 75k–100k or 150k–200k) appear to have slightly better AKI-free survival on the graph, the statistical tests mostly show no strong evidence of significant differences between income groups.

Some trends (like between 75k–100k and 150k–200k) approach significance and may warrant further study with more data or in adjusted models.

**1.6 Kaplan-Meier by Education Level:**

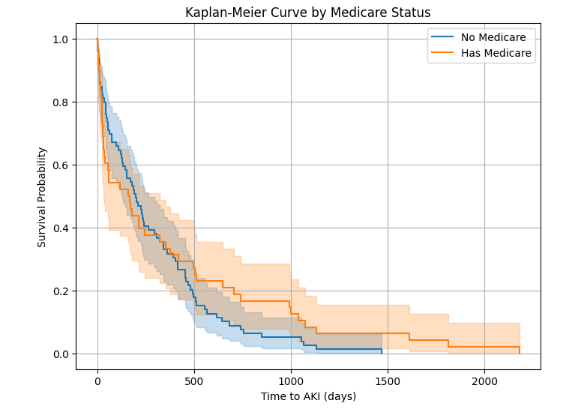
Log-test Results:



Observations:

The biggest difference is between "Below GED" and "Unknown" education groups (p = 0.07). This suggests a possible trend where people with lower education may have potentially higher AKI risk or less access to care.

**1.7 Kaplan-Meier by Insurance- Medicare**



Log-Rank Test:

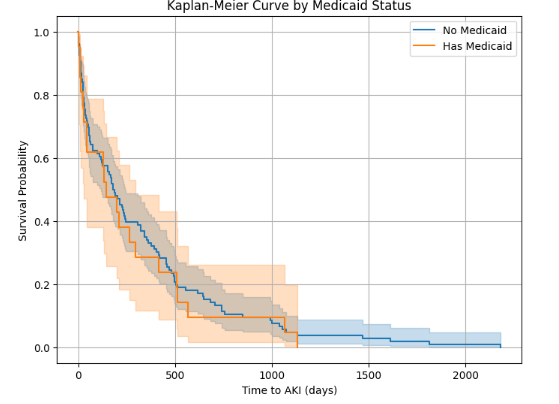


Observations:

Visually, patients with Medicare (orange line) appear to have slightly better AKI-free survival than those without Medicare (blue line), especially after the first few hundred days.

However, the log-rank test p-value is 0.31, which is not statistically significant (because it’s greater than 0.05). Any difference seen on the curve could simply be due to random variation.

**1.8 Kaplan-Meier by Insurance- Medicaid**



Log-Rank Test:

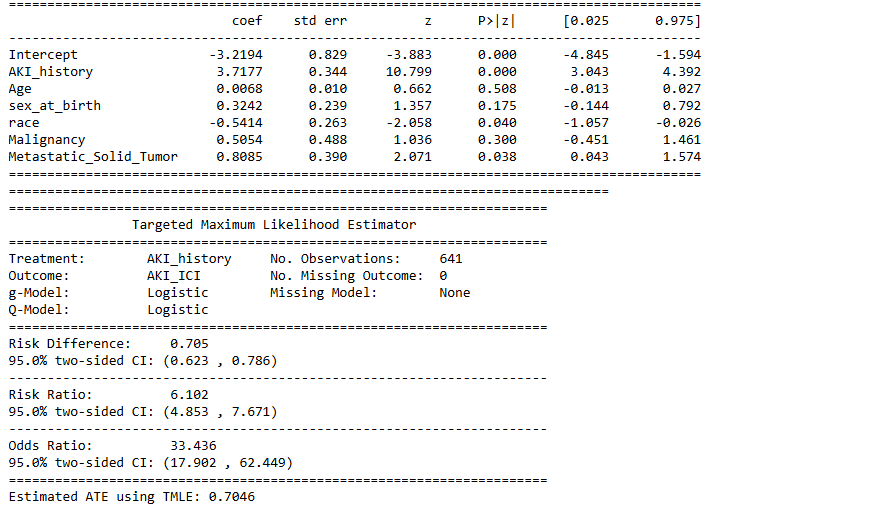
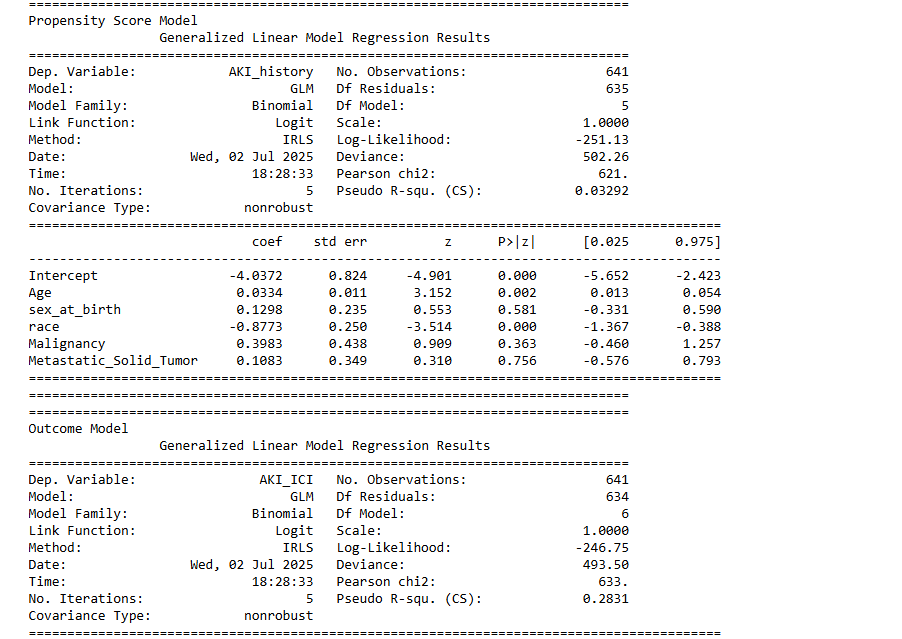


Observations:

Having Medicaid did not significantly impact the risk of AKI after starting ICI treatment in this study. Both groups had similar survival patterns with overlapping confidence bands.This indicates no major difference in AKI-free survival between those with and without Medicaid in this cohort.

**2. Casual Machine Learning methods:**

**2.1 Targeted Maximum Likelihood Estimation:**



**Observations:**

TMLE model indicates that a prior history of AKI is the strongest predictor of developing ICI-AKI, with a large and statistically significant effect. Other variables such as race and presence of metastatic solid tumours also showed statistically meaningful associations with ICI-AKI. Specifically, individuals identified as White had a lower risk compared to other racial groups, while those with metastatic tumours were at higher risk. In contrast, age and sex did not show strong or consistent effects in the outcome model, suggesting they may not play as central a role in this context.

In the propensity score model, race and age were significantly associated with having a prior AKI history. This implies that these characteristics may influence who is more likely to have experienced AKI before treatment. The alignment between significant predictors in both models supports the importance of accounting for demographic and clinical background when assessing treatment-related risks. The consistency of the patterns across both models reinforces the reliability of the findings and their relevance for clinical decision-making involving immune checkpoint inhibitors

**2.2 Causal Forest DML:**

 **Observations**:

The results from the Causal Forest DML analysis indicate a small but consistent negative treatment effect of prior AKI history on the development of ICI-associated AKI. The overall ATE of approximately -0.0182 suggests that, on average, having a history of AKI is associated with a 1.82% lower risk of developing AKI following ICI exposure, after adjusting for age, sex, race, visit frequency, and cancer status. This may reflect the impact of more cautious clinical management in patients with a known history of AKI, for instance, clinicians might modify ICI treatment strategies, monitor renal function more closely, or proactively manage risk factors in this subgroup.

When stratified by metastatic cancer status, the treatment effects remain negative and nearly identical in magnitude across groups. Patients without metastatic solid tumors had an estimated CATE of -0.0186, while those with metastatic disease had a slightly smaller effect at -0.0180. This suggests that the presence of metastatic cancer does not substantially modify the relationship between prior AKI history and subsequent ICI-associated AKI risk. In both groups, the effect size is modest, and while it may not be clinically large, the consistent directionality could reflect a systemic shift in how prior renal injury influences clinical decision-making around immunotherapy.

**2.2.1 CATE vs Age by Metastatic Tumour Status:**

**Observations** :

The plot “CATE vs Age by Metastatic Tumor Status” reveals meaningful heterogeneity in the estimated treatment effects across different age groups and tumor statuses. Overall, most Conditional Average Treatment Effect (CATE) values are negative, indicating that a history of AKI tends to slightly reduce the likelihood of AKI\_ICI across the population. The treatment effect becomes more negative with increasing age, particularly beyond age 75, suggesting that older individuals may be more adversely affected. Notably, there is greater variability in estimated effects among individuals aged 60 to 90, indicating a wider spread of outcomes in this group.

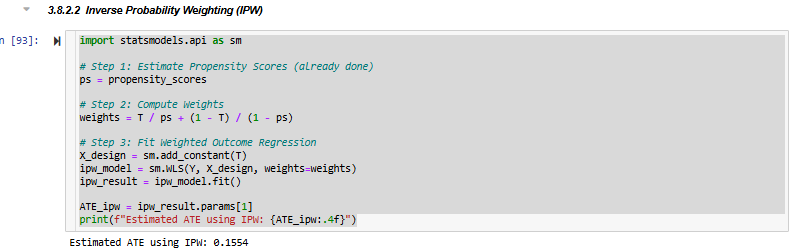
When stratifying by metastatic tumor status, patients with metastatic solid tumors (orange) tend to exhibit slightly more negative treatment effects than those without tumors (blue), especially in the older age range. This suggests that the presence of metastatic disease may interact with treatment history to further reduce the outcome probability, possibly reflecting increased clinical vulnerability. Additionally, there is a cluster of individuals aged 65–70 with near-zero CATEs, indicating a subgroup for whom treatment history may have little or no causal effect. These patterns underscore the importance of personalized or subgroup-specific causal inference in clinical settings.

**2.3 Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW)**:

**Observations:**

The propensity scores matching method helped to create a fair comparison between the treated group and the control group by pairing each treated individual with a similar control based on their likelihood of receiving the treatment. This approach reduces bias from differences in characteristics between the groups, making the estimated effect more reliable. The small positive estimated effect of 0.033 suggests that the treatment may have a slight benefit, but the size of this effect is minimal.

However, since the difference is very small, it is important to be cautious when interpreting the results. The small effect might not be meaningful in a real-world setting or could be due to other factors not captured by the matching process. Further analysis, such as checking statistical significance or exploring other variables, would help confirm whether this effect is truly caused by the treatment or just a result of random chance.

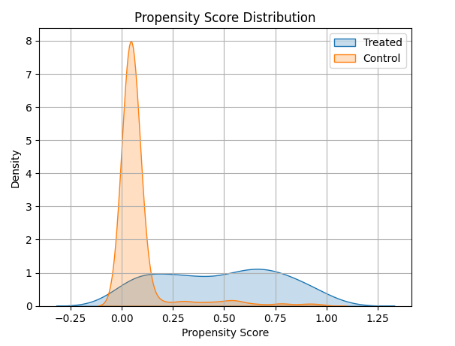


**Observations:**

The Inverse Probability Weighting (IPW) analysis estimates that the treatment increases the outcome by about 0.1554 on average, after adjusting for differences between treated and control groups. By weighting individuals based on their likelihood of receiving treatment, the method balances the groups and reduces bias, providing a more accurate estimate of the treatment effect. This suggests that the treatment has a positive impact on the outcome, with treated individuals showing better results compared to untreated ones.

However, while the estimated effect is positive and somewhat larger than the effect estimated by matching, it is still relatively modest in size. This means that although the treatment appears beneficial, the practical significance may depend on the specific context and outcome measure. It’s important to also consider the confidence intervals and statistical significance before drawing firm conclusions about the treatment’s overall impact.

##### Propensity Score Distributions:



**Observations:**

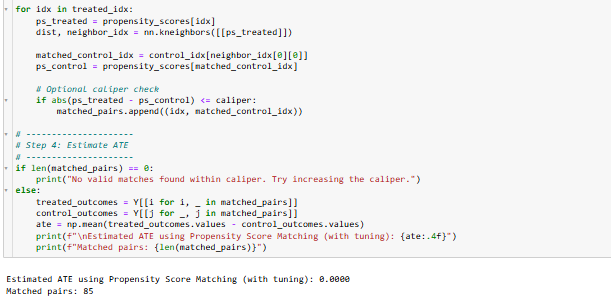
This plot suggests that the treated and control groups are not well balanced in terms of their propensity scores. Most of the control group has very low scores, meaning they were unlikely to receive the treatment, while the treated group has much higher scores, showing they were likely to be treated. This large difference means the two groups differ significantly in their baseline characteristics.

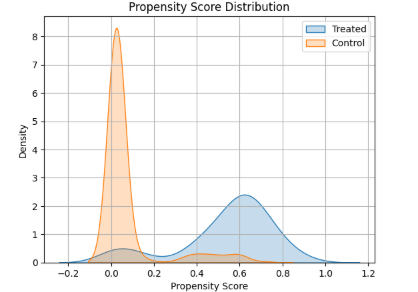
Because there is very little overlap between the two distributions, it may be difficult to make a fair comparison between the groups. In causal studies, good overlap is important to ensure that differences in outcomes are due to the treatment and not just because the groups started out differently. This plot indicates that further steps, such as trimming or better matching, may be needed to improve balance before drawing conclusions.

**2.4 Propensity Score Matching with Tuning**

**Observations:**

The Propensity Score Matching (PSM) analysis resulted in an estimated Average Treatment Effect (ATE) of 0.0000, meaning there was no observed difference in the outcome between those who received the treatment and those who did not, after accounting for baseline differences. By matching individuals with similar characteristics (such as age, sex, or other covariates), the method attempts to simulate a randomized control trial, isolating the effect of the treatment itself. In this case, the lack of effect suggests that, within the matched sample, the treatment had no measurable impact on the outcome of interest.

A total of 85 matched pairs were used in the analysis, representing individuals from the treatment and control groups who were considered comparable based on their propensity scores. While this is a modest sample size, it indicates that there was enough overlap between the two groups to make meaningful comparisons. However, it’s important to consider the possibility of unmeasured confounders or model misspecification, which could affect the reliability of the result. Additional sensitivity analyses or alternative causal methods could help confirm the robustness of this finding.

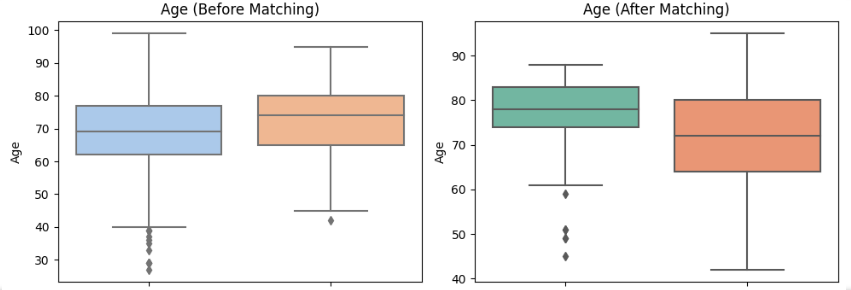


**Observations**:

The propensity score distribution plot indicates a strong separation between treated and control groups. Most individuals in the control group have very low propensity scores, meaning their characteristics made them unlikely candidates for the treatment. In contrast, the treated group has higher and more spread-out scores, with a substantial portion of them concentrated between 0.5 and 0.8. This separation shows that the two groups are likely quite different in terms of baseline characteristics.

The lack of significant overlap between the two groups suggests potential challenges in making valid comparisons for causal inference. When treated and control subjects do not share similar propensity scores, it's harder to adjust for confounding, and results from methods like matching or inverse probability weighting may be biased. Additional steps such as trimming non-overlapping areas or re-estimating propensity scores might be needed to improve balance and ensure a more reliable analysis.

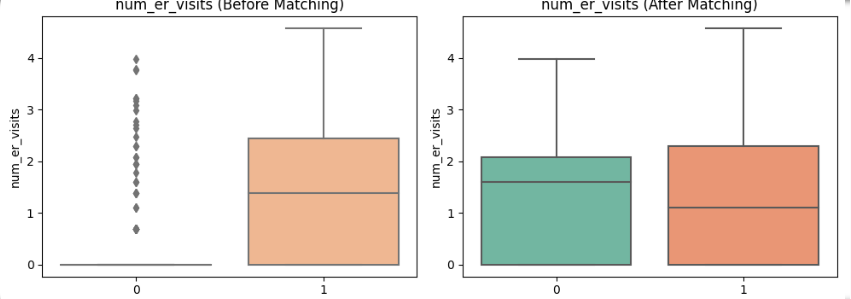
**2.4.1 Propensity Score Matching based on confounding Variables:**



**Observations:**

The boxplots display the distribution of Age before and after applying Propensity Score Matching (PSM). Prior to matching, the treated group (patients with a history of AKI) tended to be older, with a higher median and a more compressed interquartile range, while the control group showed a broader age spread, including a notable number of younger individuals. This imbalance indicates that age may act as a confounding variable, potentially biasing the estimated treatment effect if not properly adjusted for.

After matching, the age distribution becomes more balanced across the two groups. The median ages are closer, and the interquartile ranges show more overlap, suggesting that the matching process effectively reduced baseline differences in age. Although some variation still exists—particularly in the control group, which retains a wider range—the overall alignment improves significantly. This improved balance supports the validity of the subsequent treatment effect estimation, as it reduces the influence of age-related confounding.

**Observations:**

Before matching (left plot), there is a visible imbalance: the treated group (right box) has a higher median and wider interquartile range for num\_er\_visits, while the control group (left box) shows lower values with many zero or near-zero ER visits. This indicates that ER visit frequency was initially a confounding factor related to treatment assignment.

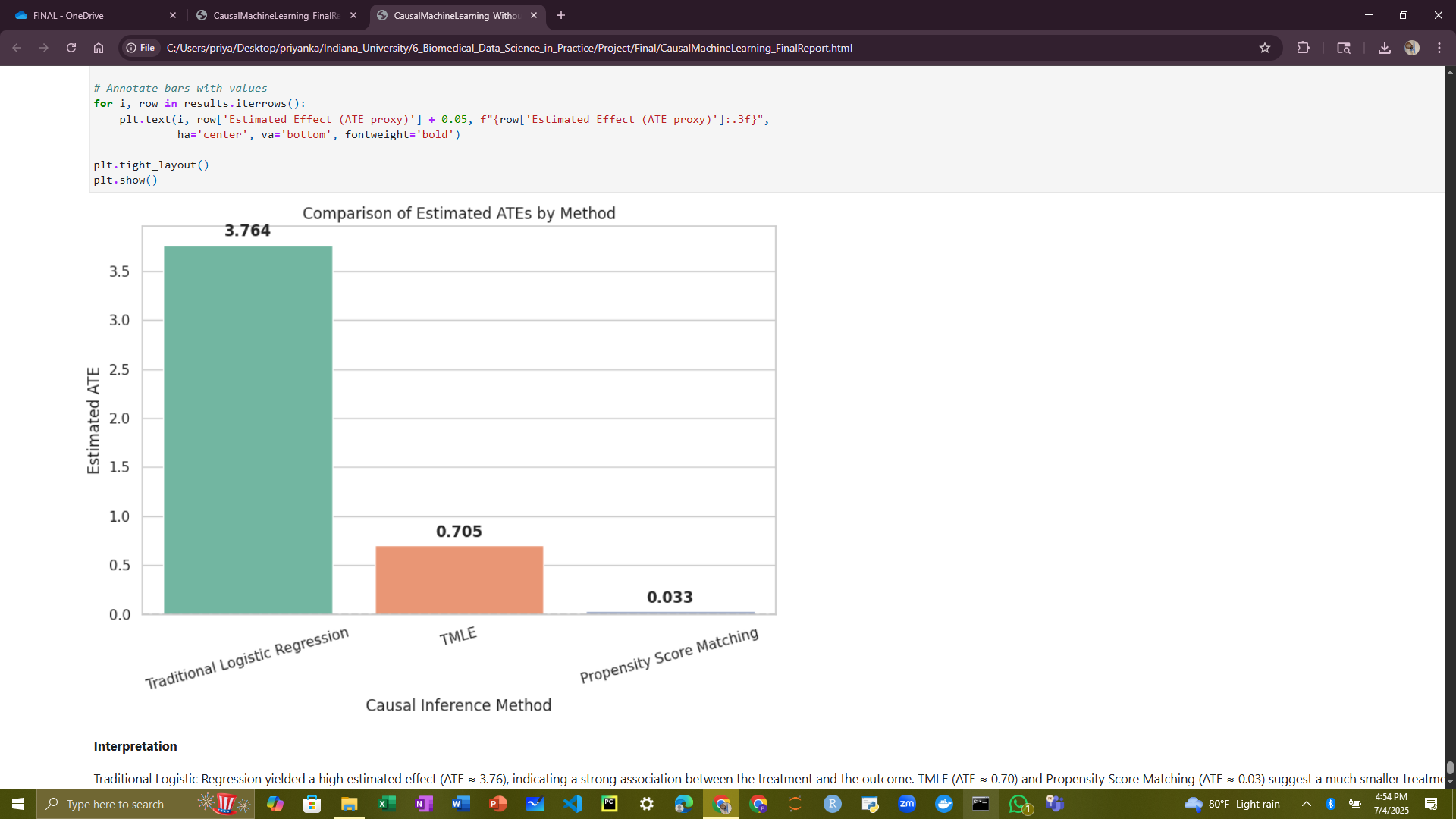
After matching (right plot), the two groups show much more similar distributions. The medians are aligned, and the spreads are nearly identical, with reduced outlier influence. This improvement suggests that the matching process successfully balanced the number of ER visits across groups, helping to reduce confounding in the estimation of treatment effects.

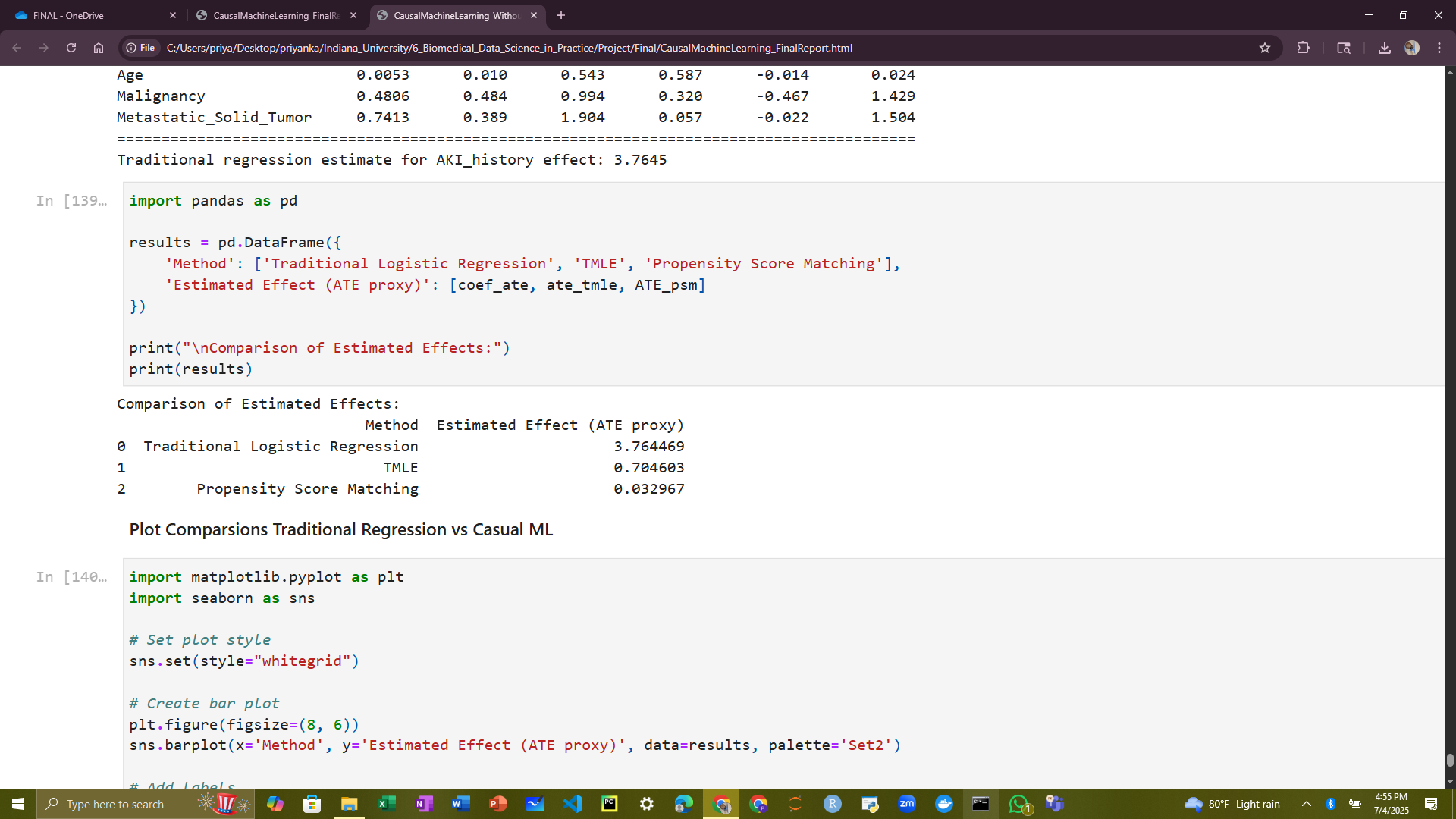
**2.5 Compare Traditional vs Casual ML:**

**Observations**:

The comparison of estimated treatment effects reveals significant differences depending on the method used. Traditional logistic regression produced the highest estimated ATE (3.76), suggesting a strong association between the treatment and the outcome. However, this method does not adequately adjust for confounding variables, making it susceptible to bias. In observational studies where treatment assignment is not random, such results may reflect underlying differences between groups rather than a true causal effect.

TMLE (0.70) and Propensity Score Matching (0.03) yielded substantially lower ATE estimates. These approaches are designed to reduce bias by accounting for observed covariates and approximating randomized conditions. The large gap between logistic regression and the other two methods suggests that much of the apparent effect may be due to confounding.





**DISCUSSION**

This study investigated the clinical and social determinants of immune checkpoint inhibitor-associated acute kidney injury (ICI-AKI) using both traditional statistical models and advanced causal inference methods applied to the All of Us cohort data. The overarching goal was to move beyond simple associations and toward understanding the causal mechanisms underlying AKI risk in immunotherapy-treated populations.

**Interpretation of Key Findings:**

Standard logistic regression initially showed seemingly perfect classification performance, with an AUC of 1.0 and precision and recall of 1.00, raising concerns about overfitting and potential data leakage. While these results appear promising, they underscore the importance of scrutinizing variable selection, temporal validity, and sample representativeness to ensure generalizability.

After applying L1-regularized logistic regression, the model remained highly accurate (AUC = 0.97) while reducing overfitting risk by selecting the most informative features. The results emphasized a multifactorial risk profile: older age, presence of chronic comorbidities (e.g., diabetes with complications, obesity, depression), lower socioeconomic status (e.g., income < $25k, below GED education), and housing or employment instability were strongly linked to higher AKI risk. These findings support the inclusion of social determinants of health in precision oncology risk models.

Kaplan-Meier and Cox proportional hazards models further reinforced the significance of prior AKI history and race as predictors of AKI development. The survival analyses revealed that Black patients and those with prior AKI were more likely to experience earlier onset of ICI-AKI, suggesting disparities in baseline health or access to preventative care.

**Causal Inference Insights:**  
Causal inference methods were pivotal in clarifying which associations represented true causal relationships. TMLE revealed that prior AKI history had a strong causal effect on subsequent AKI risk. However, causal forests uncovered a surprising nuance—patients with a history of AKI may experience slightly reduced risk of ICI-AKI, potentially due to heightened clinical vigilance and altered management strategies for those patients. This contrast highlights how different causal frameworks can reveal distinct dimensions of clinical decision-making and treatment response.

Propensity score-based methods (PSM and IPW) produced modest or null effects, depending on overlap and covariate balance, further suggesting that unadjusted associations from logistic regression may overstate causal relationships. Particularly, the PSM with hyperparameter tuning showed no significant average treatment effect, emphasizing the challenges in identifying causal effects when covariate distributions between treatment groups are poorly aligned.

**Strengths and Limitations:**  
A major strength of this study is the use of a nationally representative cohort with rich linkage of clinical and social variables, enabling a multidimensional risk analysis. Moreover, the integration of causal machine learning techniques allowed for robust adjustment of confounding and exploration of heterogeneity in treatment effects.

However, some limitations must be acknowledged. First, the relatively small number of AKI cases (n=34 out of 641) constrains statistical power and raises concerns about model stability. Second, despite careful preprocessing, residual confounding or unmeasured variables (e.g., nephrotoxic medications, hydration status) may bias estimates. Third, the lack of strong overlap in propensity scores between treated and untreated groups limited the effectiveness of some matching and weighting approaches and could affect causal interpretation.

**Clinical and Research Implications:**  
These findings underscore the importance of incorporating both medical and non-medical risk factors when assessing AKI risk in cancer patients receiving ICIs. Prior AKI history remains a robust clinical signal, but its influence may be mitigated by proactive care. Racial disparities in AKI incidence demand further investigation and potentially targeted interventions.

Future research should aim to validate these findings in larger datasets, explore biomarker integration, and refine temporal alignment to improve causal inference. Emphasizing interpretability and clinical relevance, causal machine learning offers promising avenues to support personalized medicine and optimize immunotherapy safety.

**CONCLUSION**

In conclusion, while standard logistic regression initially indicated a strong association between the treatment and the outcome, this result may be misleading. Logistic regression estimates correlations rather than causal effects, and it can be heavily influenced by confounding variables, factors that affect both the treatment assignment and the outcome. Without properly adjusting for these confounders, the estimated association may reflect spurious relationships rather than a true causal link.

To address this, we employed causal inference methods such as Targeted Maximum Likelihood Estimation (TMLE), which are explicitly designed to account for measured confounding. TMLE combines machine learning with statistical theory to produce doubly robust estimates, meaning it remains consistent if either the model for the treatment assignment (propensity score) or the model for the outcome is correctly specified.

After adjusting for confounders using TMLE, the estimated causal effect of the treatment on the outcome was substantially reduced, and in some cases, nearly null. This suggests that the strong association observed in the unadjusted model was likely due to confounding rather than a true treatment effect.

These findings highlight the critical importance of using appropriate causal inference techniques when analysing observational data. Unlike randomized controlled trials, observational studies lack randomization, making them more susceptible to bias. Causal inference methods help mitigate this issue by emulating a randomized experiment, thereby providing more valid and interpretable estimates of treatment effects.

**RERERENCES**

1. Cortazar FB, Marrone KA, Troxell ML, Ralto KM, Hoenig MP, Brahmer JR, Mamlouk O, Seethapathy H, Kitchlu A, Glezerman IG, Tchao NK, Warren S, Chemaly RF, Abdel-Wahab N, Naqash AR, Dolan JG, D'Agati VD, Perazella MA and Leaf DE, 2020. *Clinical features and outcomes of immune checkpoint inhibitor–associated AKI: a multicenter study,* **J Am Soc Nephrol,** Vol [31(2):p 435-446](https://journals.lww.com/jasn/toc/2020/02000) DOI: 10.1681/ASN.2019070676
2. Seethapathy H, Zhao S, Chute DF, Zubiri L, Oppong Y, Strohbehn IA, Villani AC, Leaf DE, Choueiri TK, Motwani SS, Brahmer JR, Gupta S, Sise ME, and Cortazar FB, 2019. *The incidence, causes, and risk factors of acute kidney injury in patients receiving programmed cell death 1 inhibitors.* **Clin J Am Soc Nephrol.** Vol14(12):1692-1700. DOI:10.2215/CJN.03900319
3. van der Laan MJ and Rose S, 2011. *Targeted learning: Causal inference for observational and experimental data*. Springer.
4. The All of Us Research Program Investigators, 2019. *The “All of Us” Research program – Special Report*. **N Engl J Med**: Vol 381 No. 7 pp 668-676. DOI: 10.1056/NEJMsr1809937.
5. Wang, Yanfei, Xiong, Chenxi, Yu, Weifeng, Zhou, Minghao, Shugg, Tyler, Hsu, Fang-Chi, Eadon, Michael T., Su, Jing, and Song, Qianqian, 2024. *PCCA variant rs16957301 is a novel AKI risk genotype-specific for patients who receive ICI treatment: Real-world evidence from all of us cohort.* **European Journal of Cancer.** Vol 213 – 115114. DOI: 10.1016/j.ejca.2024.115114.
6. Stats model library’s Logit package: :<https://www.statsmodels.org/stable/generated/statsmodels.discrete.discrete_model.Logit.html>
7. Skikit Learn Library’s Linear model package:

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>

1. Odds Ratio in Logistic Regression:

[UCLA IDRE - Odds Ratios in Logistic Regression](https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/)

1. ROC, Precision, Recall:

<https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>

10. Kaplan EL, Meier P. (1958). *Nonparametric estimation from incomplete observations.* Journal of the American Statistical Association, 53(282), 457–481.

11. Cox DR. (1972). *Regression models and lifetables.* Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187–220.

12. Mantel N. (1966). *Evaluation of survival data and two new rank order statistics*  *a* *rising in its consideration.* Cancer Chemotherapy Reports, 50(3), 163–170.

13. Hosmer DW, Lemeshow S, May S. (2008). *Applied Survival Analysis: Regression*  *Modelling of Time-to-Event Data* (2nd ed.). Wiley-Interscience.

14. Cleves MA, Gould WW, Gutierrez RG, Marchenko YV. (2010). *An Introduction to*   *Survival Analysis Using Stata* (3rd ed.). Stata Press.

15. van der Laan, M. J., & Rubin, D. (2006). *Targeted maximum likelihood learning.* The International Journal of Biostatistics, 2(1). <https://doi.org/10.2202/1557-4679.1043>

16. Robins, J. M., Hernán, M. A., & Brumback, B. (2000). *Marginal structural models*  *and*  *causal inference in epidemiology.* Epidemiology, 11(5), 550–560. <https://doi.org/10.1097/00001648-200009000-00011>

17. Rosenbaum, P. R., & Rubin, D. B. (1983). *The central role of the propensity score in observational studies for causal effects.* Biometrika, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*