In this document, we will analyze the results of the prompts that we gave to different AI models for the assignment (ChatGPT 5.1 and Claude Sonnet 4.5). Below are only some examples of the generated code. After the notebook we will also look into the report we prompted. In the end, we also tried to get the opinion of a different model to the one that generated the code (Opinion of ChatGPT 5.1 of Claude Sonnet 4.5 code).

1. Setup and Data loading

The code successfully loaded all the provided data sets to the notebook, and all libraries were imported successfully.

1.3. Dataset 1: Membership data (2020-2022)

The membership data was imported successfully, and the data is correct. This step was also done successfully in cells (1.4., 1.5.)

2. Summary of Data Structure

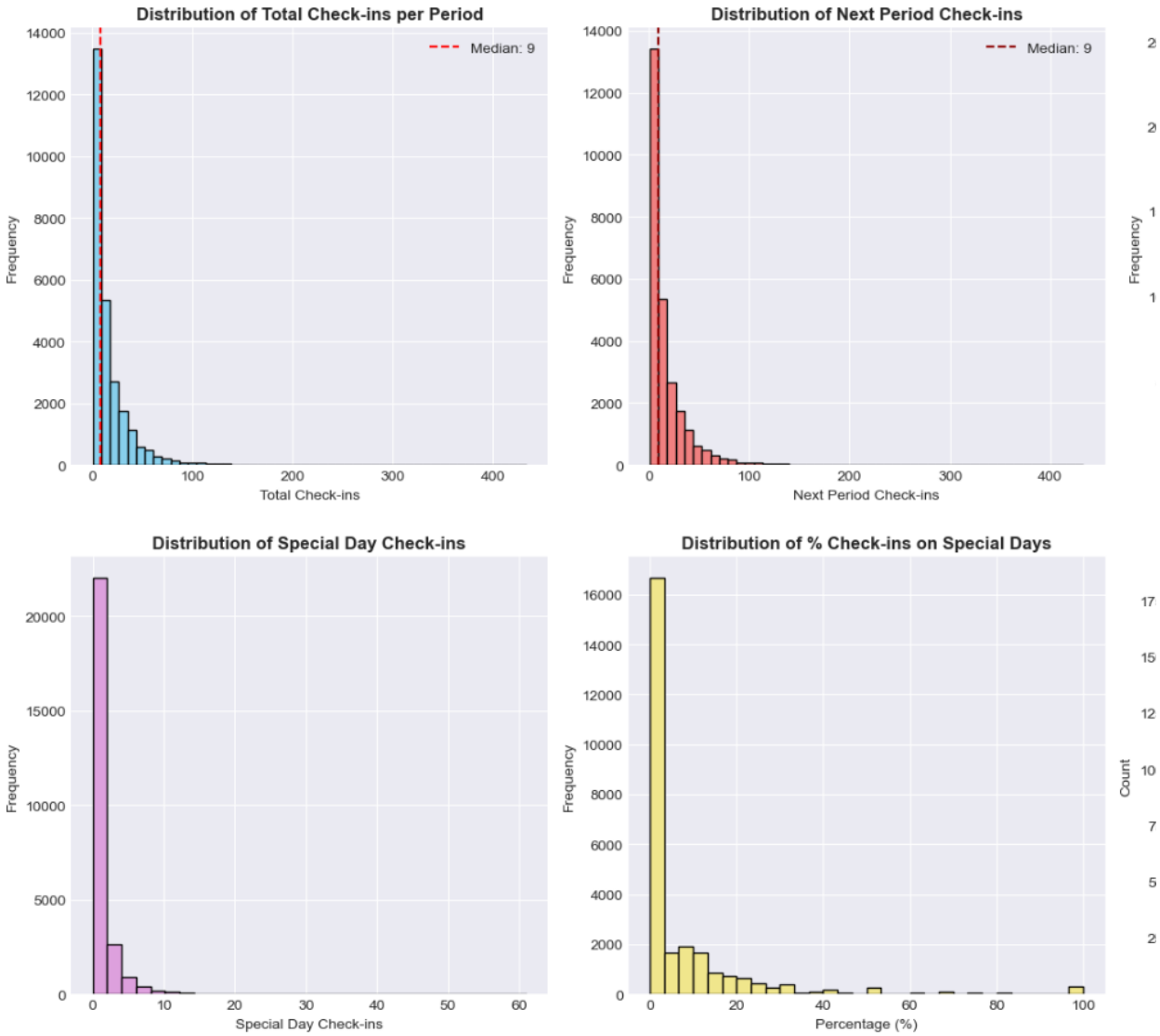
The code has found good key points from the data and talked about them well. In the Target Variables part the model has used the option that were provided in prompt number 4(Notebook_prompts file).

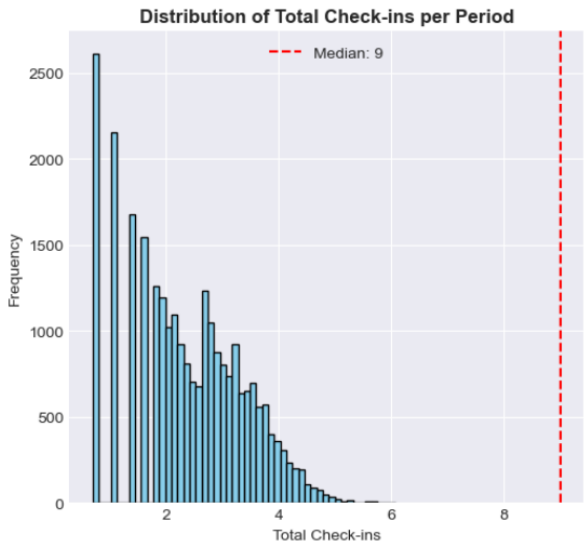
3. Data understanding and column analysis

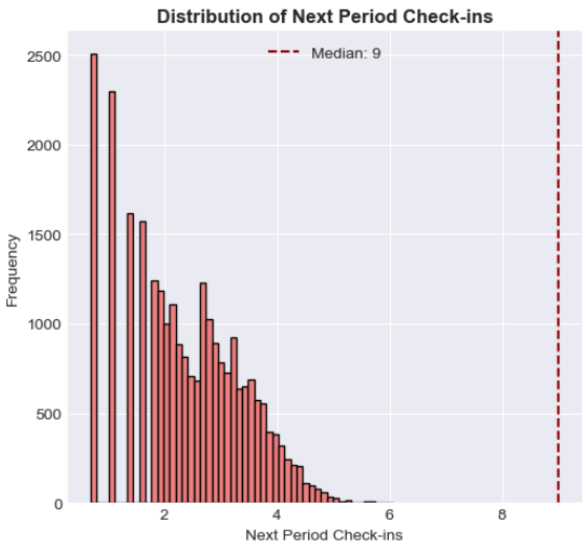
In this part, the data has been classified for the next steps. I think that the classification has been successful.

7. Descriptive Statistics and Correlations

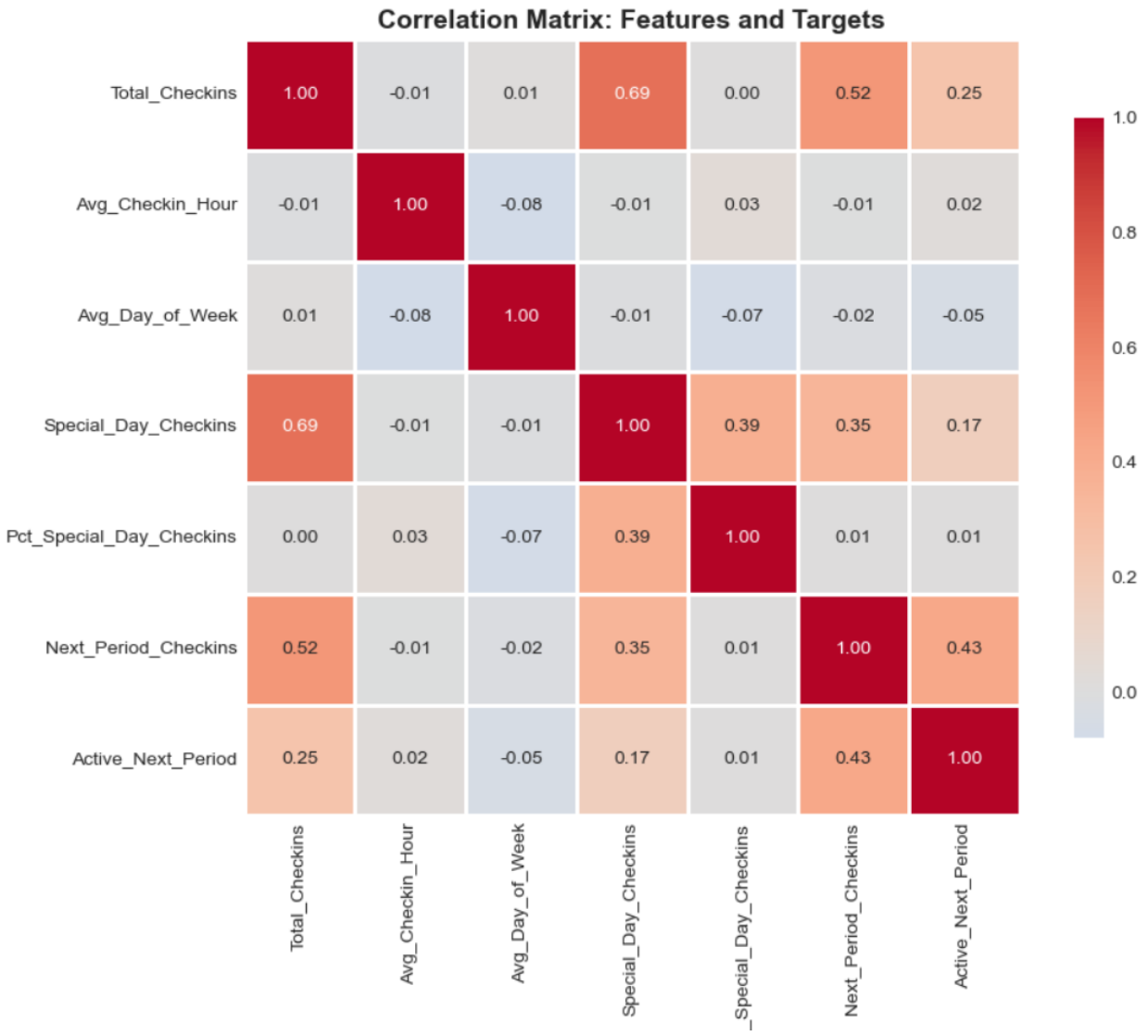
The model and the given prompt also did not give out fully sensible answers as can be seen in the picture below. The graphs below are very heavily right-skewed. I got the first two fixed by asking ChatGTP 5.1 to make a fix for me. It recommended me to add np.lo1p to the code (This can be seen in commit “log added”)




Distribution of Total Check-ins per Period
Median: 9
Frequency
Total Check-ins
0
500
1000
1500
2000
2500
2
4
6
8


Distribution of Next Period Check-ins
Median: 9
Frequency
Next Period Check-ins
0
500
1000
1500
2000
2500
2
4
6
8

The correlation matrix was also working well and was really well readable. It does not have as many data points as the one we made for the project previously.



The examples above were just some parts of the notebook, but personally I think that the model (Claude Sonnet 4.5) did a fine job with the prompt provided by ChatGPT 5.1. The whole notebook is not perfect with some issues, for example the graphs above, but as for most of the notebooks it is functional and it displays the data in many ways. For only four prompts provided by the user, this notebook is pretty good. The notebook could most likely be made even better with more time spent with the prompting.

I was also interested in asking a model for its opinion on the final product. We provided the whole project to ChatGPT 5.1 and let it make a summary of the issues that it found in the notebook. Below is the list of things that the model found. All of the points are not totally accurate, for example the first point does not cause an error in the running of the code. But also, there are good points, for example **Limited discussion of caveats** is good to consider.

- **Ridge model bug**
  You pass `random_state=42` to `Ridge`, which doesn't support that argument and will cause an error.
- **Weak justification for the target**
  The cutoff `Next_Period_Checkins ≥ 5` for "active" is arbitrary and not clearly motivated by business logic or data.
- **No time-aware validation**
  You use a random `train_test_split` even though this is a time-based prediction problem, which can give overly optimistic performance.
- **DAG not connected to modeling choices**
  The causal graph is nicely drawn but not used to justify which variables are included/excluded or how bias is handled.
- **Important features not used**
  Variables like membership type, subscription, and affiliation are discussed as important but are missing from the final feature set.
- **Crude encoding of day/hour patterns**
  Day-of-week and hour are reduced to simple averages, which poorly represent actual training patterns for circular/time-of-day data.
- **Misleading log-transformed plots**
  You log-transform check-ins but keep axes and median lines labeled/positioned as if they were on the original scale.
- **Underexploited "special days" data**
  Special days are collapsed into simple counts/percentages, without distinguishing exams, holidays, intro weeks, etc., despite that being central in your story.
- **Class imbalance and metrics not tied to the goal**
  There's no clear handling of imbalance or explanation of which metric matters most for spotting at-risk (inactive) members.

- **Limited discussion of caveats**
  You don't really discuss limitations like COVID periods, first-semester members, or missing historical data and how they affect conclusions.
- **Reproducibility assumptions**
  The notebook relies on specific folder structure and library versions without documenting them, which can hurt reproducibility.