1. DATA CLEANING

* We began by consolidating our data from three separate CSV files, namely the Applicants, Sessions, and Events tables. We addressed any potential data redundancies by removing duplicate rows, focusing primarily on the Events table, which resulted in a reduction from 9388 to 3999 rows.
* Next, we performed a thorough check of the data types for each column to ensure consistency and accuracy in our analysis.
* We merged the tables based on common identifiers, specific recruiter the session\_id and applicant\_id columns, to gain a comprehensive view of the data. Although we opted for Python for this task, SQL could serve as an alternative method.
* We streamlined our dataset by removing non-relevant columns such as first\_name and last\_name, relying instead on the applicant\_id when necessary.

1. DATA ANALYSIS

Now, let's delve into our analysis, focusing on shortening the time between two critical actions:

* Test underwriting completed by applicant
* Form reporting on the test results submitted by the recruiter

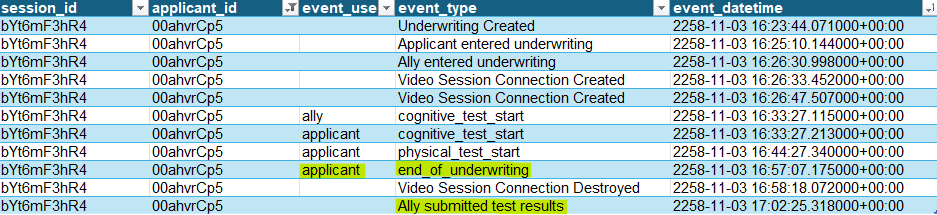
We identified these events within the event\_type column of the Events table as:

- “end\_of\_underwriting”

- “submitted test results”

We have to pay attention that the event user for the “end\_of\_underwriting” is an applicant and not an Recruiter.

Below are the relevant actions for a sampled session:



**Specific scenarios:**

* Multiple underwriting/test submissions: In such cases, we considered only the last action. For example, if there were multiple entries for the end of underwriting, we retained the final one: 
* Multiple sessions for the same applicant: Each session was treated as a distinct entity.
* Applicants who entered underwriting but did not complete it (no corresponding event in the session) yet still had the recruiter submit test results. At present, we opted to overlook such sessions as they did not meet our defined requirements.

**Submission time average and median comparison:**

Our next step will involve calculating the average and median time before and after the modification date. Simultaneously, we will address outliers. Preliminary findings indicate that the process takes minutes, thus values exceeding 2 hours will be excluded. Negative values will be discarded.

Average Time before modification: **05:09**

Average Time after modification: **12:47**

Median Time before modification: **03:47**

Median Time after modification: **06:51**

These results indicate a significant increase in the average time after the change implementation, suggesting a potential negative impact. While the median shows better results than the mean, it's still twice as high as before the modification.

**Submission time considering different parameters:**

Now we'll explore whether these results correlate with differences in dataset parameters before and after the modifications.

Age of the applicants before and after:

Average age for 'Before': **68.39** years

Average age for 'After': **67.65** years

We obtained very close results before and after the modification (68.4 vs 67.7) for this parameter. Therefore, it appears that this parameter cannot account for the increase in submission time after the change.

Submission time per gender:

Gender counts before change:

Male: **21**

Female: **36**

Gender counts after change':

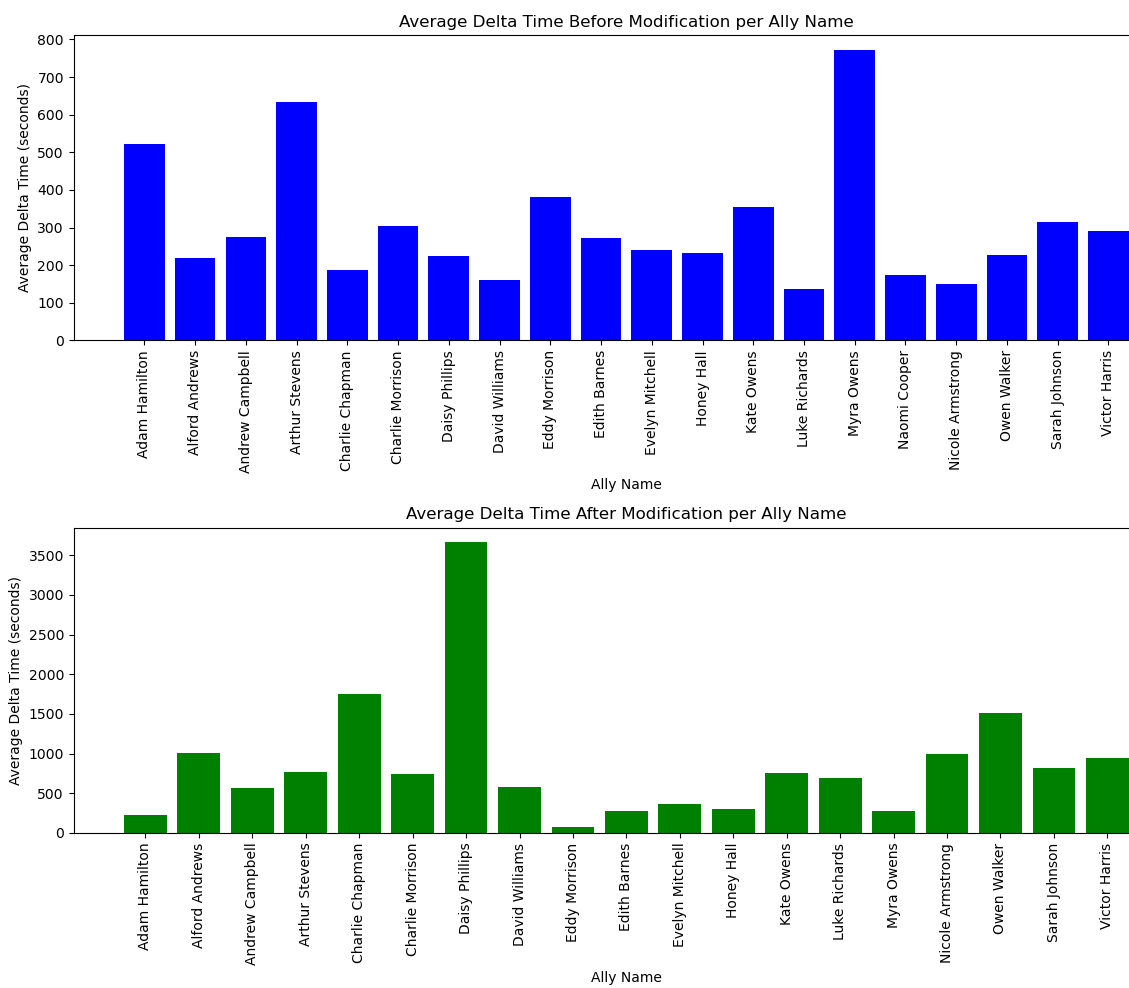
Male: **23**

Female: **42**

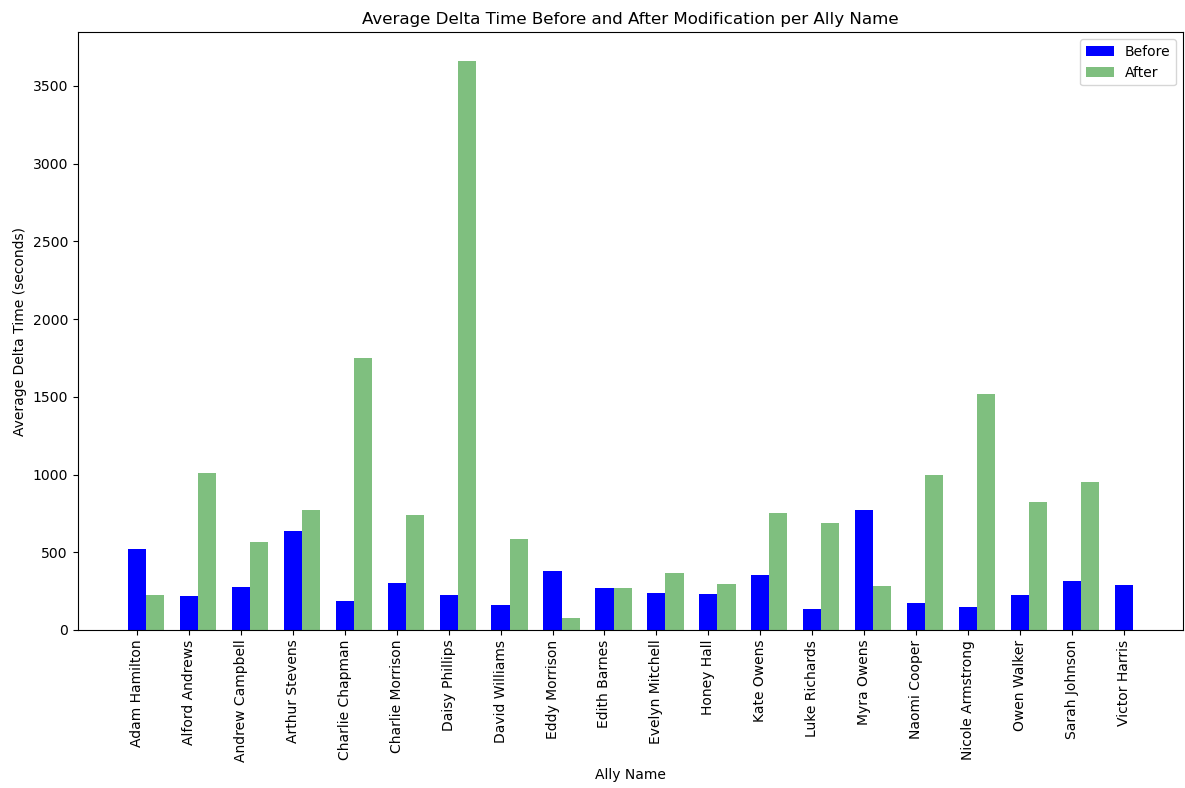
Here again, no significant difference in the gender distribution.

Submission time per Recruiter:

To ensure clarity in visualization, we will represent the submission time per recruiter using a plot.



In the same plot:



We can observe that the time after the change remains greater than before, with notably significant differences for Allies such as Daisy Phillips and Charlie Chapman.

1. CONCLUSIONS

The analysis conducted on the impact of the underwriting flow modification on the recruiter side reveals several key insights. Despite the initial expectation of a reduction in the time taken to fill in the test results form, the findings indicate an unexpected increase in both the average and median submission times following the implementation of the change.

The average time taken for submission increased from 05:09 before the modification to 12:47 after, while the median time rose from 03:47 to 06:51. These results suggest a significant deviation from the intended outcome, indicating inefficiencies introduced by the modification.

Further exploration into demographic parameters such as applicant age and gender distribution did not reveal any substantial differences before and after the modification. In addition, the examination of submission times per recruiter highlighted notable variances, with some allies experiencing more significant increases in submission time compared to others.

In conclusion, while the underwriting flow modification aimed to streamline the submission process, the analysis suggests a contrary outcome with increased submission times observed across the dataset.

Further investigation into the root causes of these delays, including the impact of recruiter training and system adaptation, is warranted to refine the underwriting process and ensure optimal efficiency in the recruiter-side submission of test results forms.