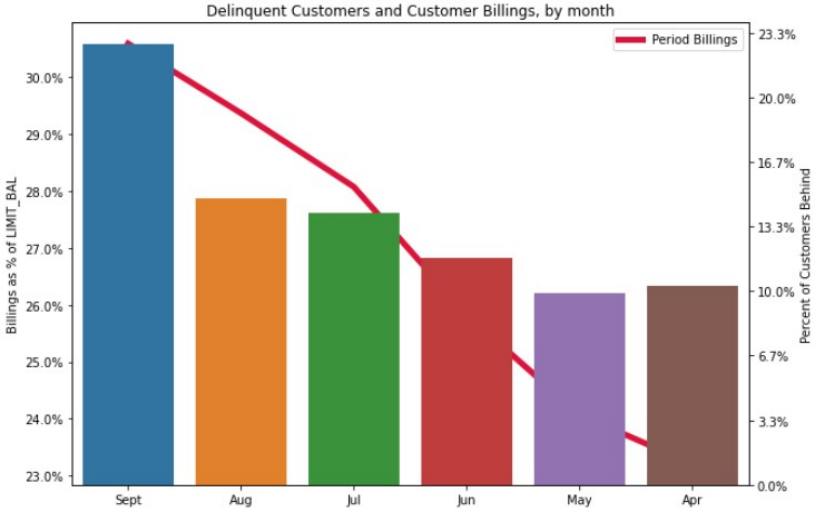
C2T2 “Lessons Learned” Report Submission: E. Jarrett (Full-time)

Key Company Takeaways (“Potential Business Value”)

- The financial picture of both Credit One, and its customers ,worsened significantly during the April – September time period
    - Aggregate monthly billing by all customers increased almost 32%, while payments made grew only 8.6%; fortunately September’s bills aren’t due for another month (not shown)
    - Gross monthly billings, as a share of total available LIMIT_BAL , grew steadily to over 30% by Sept (fig)
    - The share of customers behind in their payments more than doubled from 10% in Apr to almost 25% by Sept (fig)
    - 

    -

Primary Individual Lesson(s) Learned:

- I need to read through the *entire lesson first* to see what concepts are covered at each checkpoint. I’d dug into fairly detailed EDA already during Task 1 ‘s assignment and began fiddling with basic (default parameters) classifiers again here for Task 2 (thinking might aid visualization); lot of experimenting and likely wasted effort on my own.
- Jupyter organization is key, especially in context of above. I think I do a decent job of commenting, but now whilst preparing this writeup, I’m spending significant time/effort ‘tidying the notebook’. i.e., removing ‘orphan’ chunks of code I’d experimented with but never gotten working, or pasted functions from others’ online, mostly related to multi-collinearity checks.
- Same goes for dataframe handling – I am unsure if there’s any need/value in working with reduced column dataframes (like my ‘Summary’), or if doing so poses risks in data integrity.

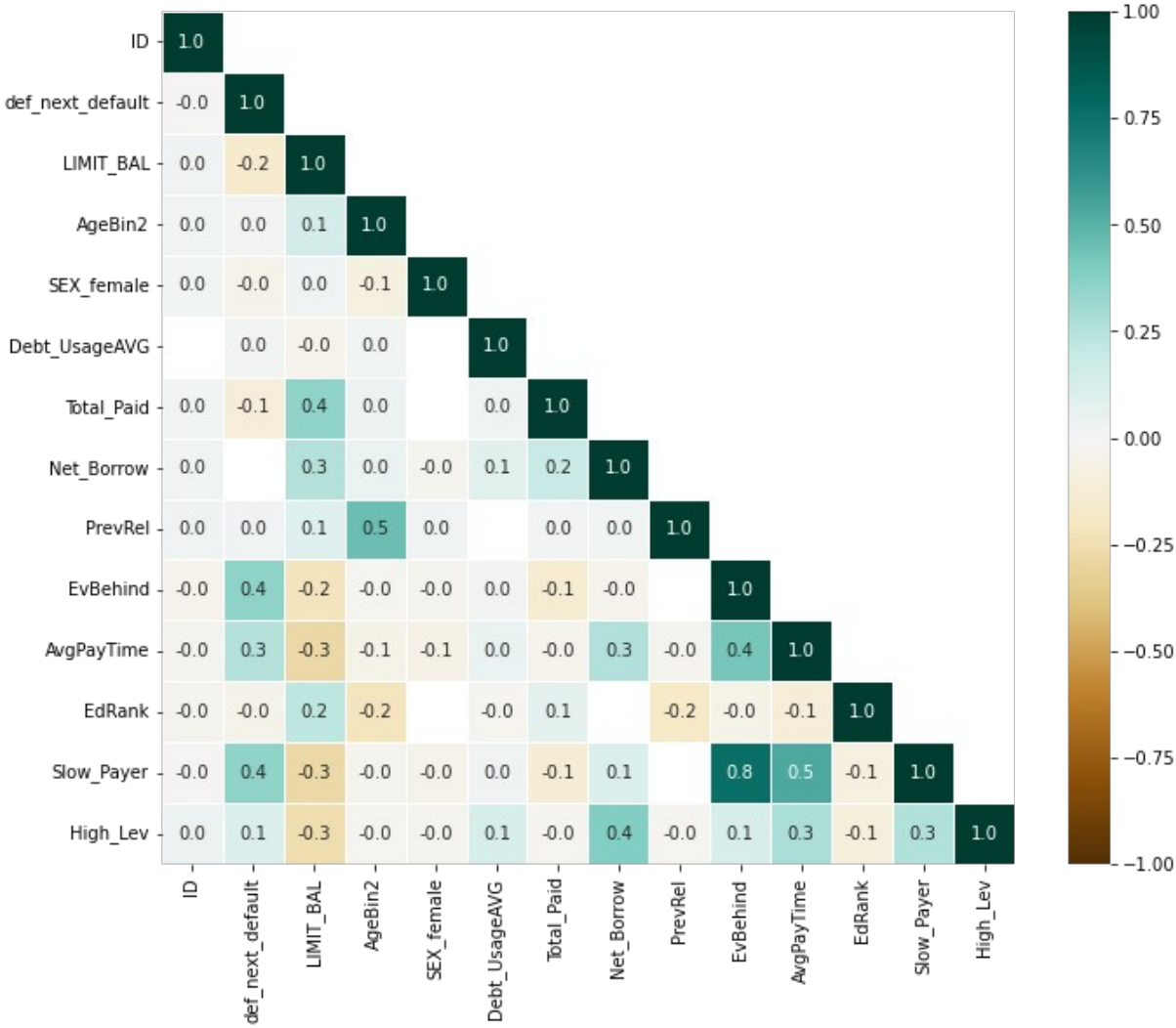
Recommendations to DS Team:

Based on preliminary analysis thus far, steady/on-time payment history is most important factor of customer creditworthiness, while demographic features are surprisingly irrelevant.

Addendum Chart Outputs:

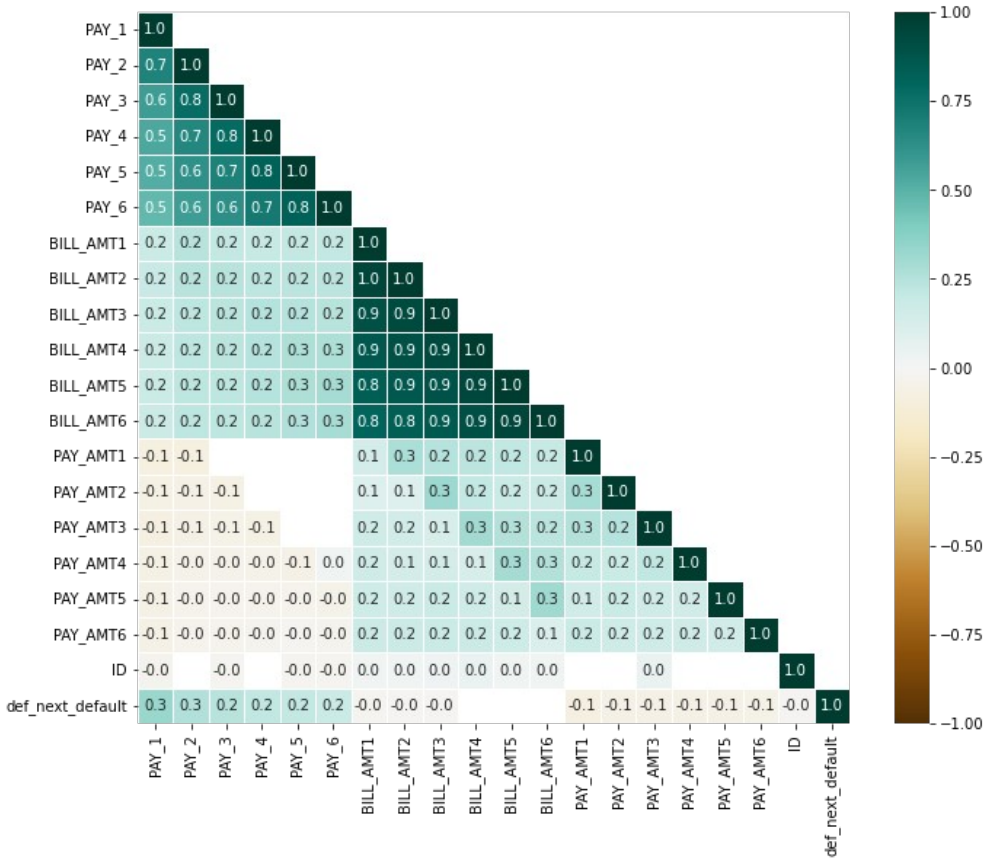
1 – Correlation Heatmap of 'Summary' Dataframe

- Multi-collinearity VIF threshold 5.0
- Masked cells (white-out) are NOT statistically significant at p_value .05



2 – Correlation Heatmap of Apr-Sept related variables

- Same statistical mask applied
- Note the diagonal-right band of 0.3 correlation starting with PAY_AMT1:BILL_AMT2 near the center… This suggests customers typically pay on lagging basis (as real world), although there are also customers who pay before statement periods end, (also real world).



3. Multi-grid of different types and scales

- Haven’t quite gotten the hang of Facet Grids and how to tackle problems with X, Y, Hue, Co l