✨ Overview

In Sprint #3 my goal was to analysing subscription data for streaming service going through three critical datasets — Cohort, A/B test and CLV. As a final output observations, conclusions on opportunities and recommendations for performace to the Product team for three separate tasks (Task 1: Cohort, Task 2: A/B Testing, Task 3: CLV) are delivered.

Steps were taken:

- **Data Cleaning**

All of the following were handled:

- Dates were properly formatted.
- Missing subscription_end values were treated as "still active" (filled with today's date).
- Duplicates, missing critical values, and rows with negative durations were removed.

- **Customers Grouping into weekly Cohorts,** showing when each Cohort start and when Cohort ends.

- The start / end of the cohort week. *Wrong outcome was corrected to weekly cohorts, which starts on Sunday and end on Saturday.*
- The number of users who subscribed that week
- The latest week any user from that cohort was active

cohort_start	cohort_end	cohort_size
2020-11-01	2020-11-07	97
2020-11-08	2020-11-14	88
2020-11-15	2020-11-21	95
2020-11-22	2020-11-28	116
2020-11-29	2020-12-05	92
2020-12-06	2020-12-12	91
2020-12-13	2020-12-19	21

- **Calculation of the total number of retained customers per cohort per week for the seven weeks** from w_1 to w_6, starting from week w_0 (active subscribers minus lost subscribers).

cohort_week_start	cohort_week_end	W_0	W_1	W_2	W_3	W_4	W_5	W_6
2020-10-26	2020-11-01	97	97	94	92	90	85	80
2020-11-02	2020-11-08	88	88	86	83	79	74	70
2020-11-09	2020-11-15	95	95	90	86	81	75	75
2020-11-16	2020-11-22	116	116	109	100	87	87	87
2020-11-23	2020-11-29	92	92	87	80	80	80	80
2020-11-30	2020-12-06	91	91	81	81	81	81	81
2020-12-07	2020-12-13	21	21	21	21	21	21	21

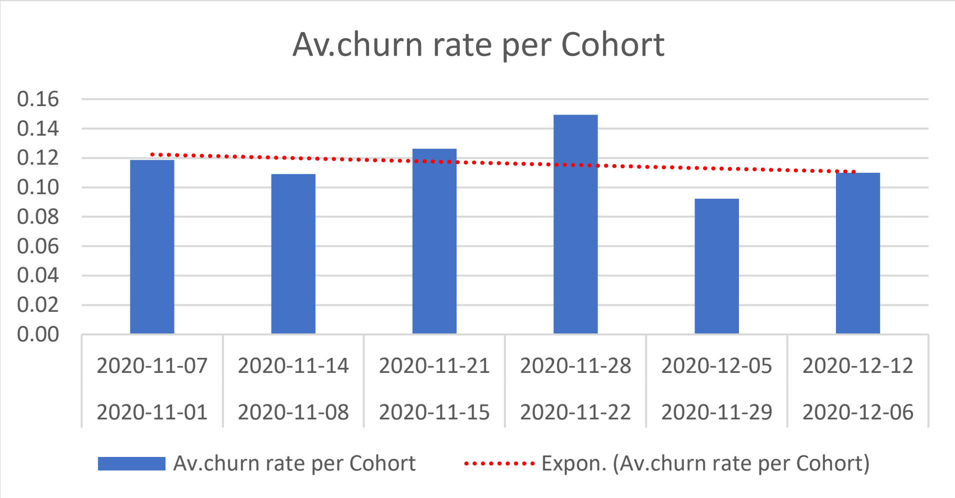
- Wrong outcome (Cohort start / Cohort end dates) was corrected to weekly cohorts, which starts on Sunday and end on Saturday.
- Data for w_1 – w_6 cohorts whose full 7 weeks haven't passed were masked, considering if today were 2020-12-20, only showing W_0. After very many attempts, right matrix was visualised:

cohort_week_start	cohort_week_end	W_0	W_1	W_2	W_3	W_4	W_5	W_6
2020-11-01	2020-11-07	97	94	92	90	85	80	72
2020-11-08	2020-11-14	88	86	83	79	74	70	0
2020-11-15	2020-11-21	95	90	86	81	75	0	0
2020-11-22	2020-11-28	116	109	100	87	0	0	0
2020-11-29	2020-12-05	92	87	80	0	0	0	0

2020-12-06	2020-12-12	91	81	0	0	0	0	0
2020-12-13	2020-12-19	21	0	0	0	0	0	0

- **Average Churn rate by weekly cohort** calculated. The churn rate reflects the percentage of users who ended their subscription within the same week they started.

cohort_week_start	cohort_week_end	Av.churn rate per Cohort
2020-11-01	2020-11-07	0.12
2020-11-08	2020-11-14	0.11
2020-11-15	2020-11-21	0.13
2020-11-22	2020-11-28	0.15
2020-11-29	2020-12-05	0.09
2020-12-06	2020-12-12	0.11
2020-12-13	2020-12-19	



Key conclusions based on the cohort and churn analysis:

📊 **1. Churn Rates Vary Across Cohorts**

- **Churn is not consistent** — it ranges from around **9% to 15%** across different weeks.
- This variability suggests that **external factors (e.g. marketing campaigns, seasonal trends, user experience changes)** may have impacted retention from week to week. Quick SWOT analysis may help.

📅 **2. Later Cohorts Show Lower Churn (e.g. Nov 29 Cohort at 9%)**

- This could imply:
  - Improved onboarding or value proposition
  - Better-targeted users during that week
  - A price promo campaign that positively affected user behavior

📈 **3. Earlier Cohorts Had a bit Higher Churn, one had very high (e.g. Nov 22 at 15%)**

- This could imply:
  - Less effective engagement strategies at the time
  - Possibly lower-quality leads or poorly matched user expectations

📌 **4. Retention is Improving**

- With the drop in churn in later weeks, there's a **positive trend in user retention**.
- This can be a **strong signal of business improvement** if the trend continues.

💡 **Strategic Suggestions:**

- Too little data to make a matured conclussions, bur few we may:
  - Investigate what changed during the **Nov 29** week that led to better retention — replicate successful actions.
  - Analyze why **Nov 15 and Nov 22** users churned — survey of exit reasons, feature usage, support tickets can help.
  - Continue monitoring cohorts weekly to validate that retention stays strong or improves.
  - Monitor changes across Cohorts_by_Device_Type, or/and merge with Cohorts_by_Country:

cohort_week	category	Users
2020-10-26	desktop	56
2020-10-26	Mobile	38
2020-10-26	Tablet	3
2020-11-02	desktop	57
2020-11-02	Mobile	28
2020-11-02	Tablet	3
2020-11-09	desktop	55
2020-11-09	Mobile	35
2020-11-09	Tablet	5
2020-11-16	desktop	70
2020-11-16	Mobile	43
2020-11-16	Tablet	3
2020-11-23	desktop	54
2020-11-23	Mobile	38
2020-11-30	desktop	48
2020-11-30	Mobile	43
2020-12-07	desktop	14
2020-12-07	Mobile	7

- **Average Retention rate by weekly cohort** calculated. The **average retention rate** tells what percentage of users, on average, remain active **each week** after they first subscribe.

cohort_week_start	cohort_week_end	Av.retention rate per Cohort
2020-11-01	2020-11-07	0.88
2020-11-08	2020-11-14	0.89
2020-11-15	2020-11-21	0.87
2020-11-22	2020-11-28	0.85
2020-11-29	2020-12-05	0.91
2020-12-06	2020-12-12	0.89
2020-12-13	2020-12-19	

Key conclusions based on the **Retention** analysis:

📈 **Trend Observed**

- Too short data we have, neverthless this suggests a **positive upward trend in retention** over observed period.

🧠 **Why Retention May Be Increasing**

1. Active Users Skewing the Recent Cohorts

- Users who subscribed later (e.g. Nov 29 or Dec 06) have had **less time to churn**.
- For example, someone who subscribed on Dec 13 is counted as **"active"** by default if their subscription_end is blank.
- Therefore, retention appears **artificially higher** in recent weeks due to **shorter observation windows**.

2. Product or Service Improvements

- If the business introduced:
  - Better onboarding
  - New features
  - Faster support
  - Targeted user education
- …these could help **new users find value faster**, reducing early churn.

3. Marketing Optimization

- Later cohorts may have come from **refined marketing campaigns**, resulting in:

    - Better audience targeting
    - Higher-quality users more likely to stay

4. Seasonality or External Events

- The time window covers **late Q4**:
    - Black Friday and Cristmas holiday preparation
    - These often bring **motivated users**, especially in subscriptions or digital products.
- These users may be **more likely to remain active**, improving retention.

To make a **more accurate conclusion**, we'd need **longer-term data** — especially post-subscription for later cohorts.

Link to conversation with ChatGPT
https://chatgpt.com/c/683ec294-75a4-8011-b014-506570918508