# Methods:

## 1. Cohorts analysis

### 1.1 Introduction

Cohorts analysis is an analytical technique that focuses on analyzing the behavior of a group of users over time, thereby uncovering insights about the experiences of those customers, and what companies can do to better those experiences. With a defined time-span, users are split into groups(cohorts) for analysis. It's usually completed by data visualization which better allows the company to realize that if the number of users (or even users in specific situation) are falling as the time period increases.

### 1.2 Results

I performed Cohorts analysis on all new users, new users with early win (win = True), and new users with early lose (win = False). The below Fig.1, Fig.2 and Fig.3 show the corresponding Cohorts metric.

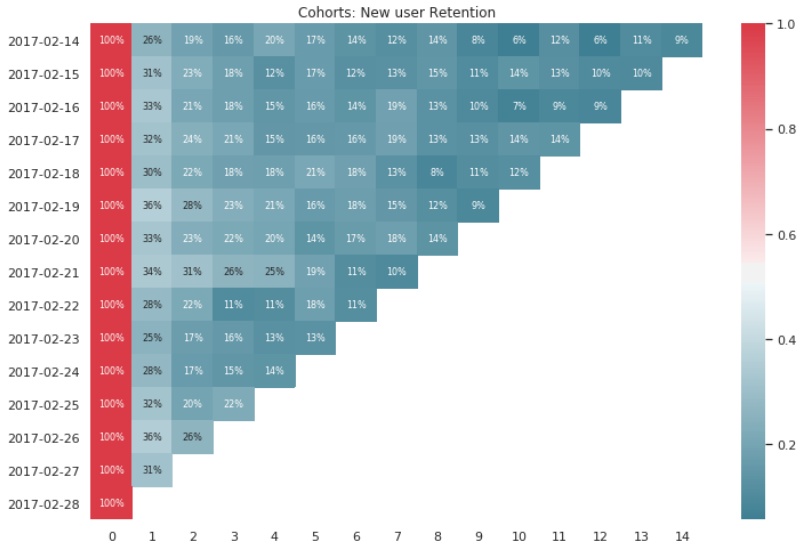


**Fig1.** New user retention cohorts

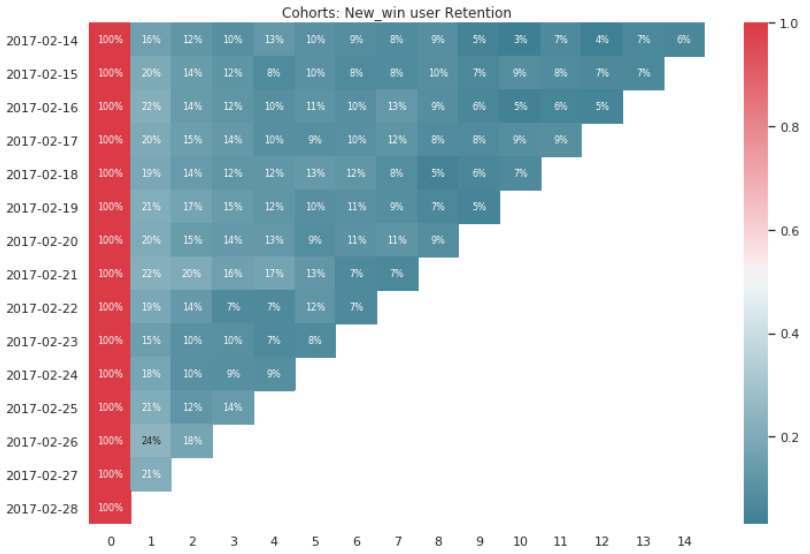


**Fig2.** New user with early win retention cohorts

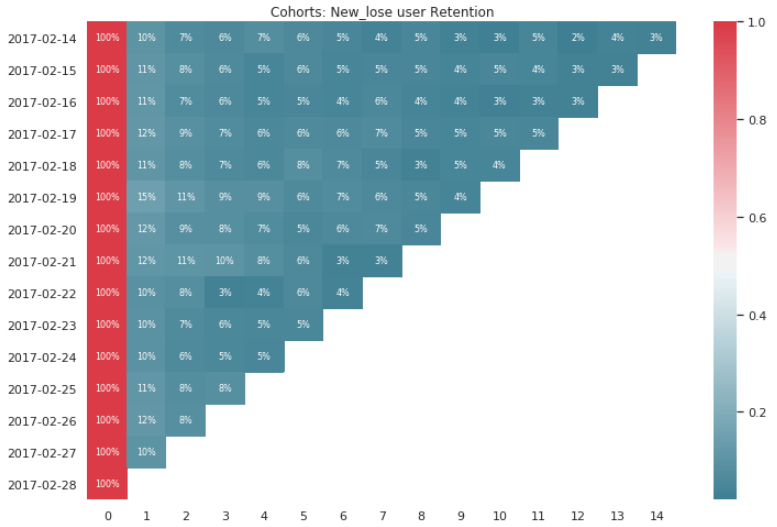


**Fig3.** New user with early lose retention cohorts

From cell to cell in **Fig.2** and **Fig.3**, it is obvious that new users with early win always have better retention value than users with early lose.

## 2. Student's t-test

2.1 Introduction

Student's t-test is a statistical method of testing hypotheses that two independent random samples have the same mean value. It tells us how significant the differences between groups are, or if those differences could have happened by chance.

A p-value of t-test is the probability that the results from sample data occurred by chance. Lower p-values always indicate your data did not occur by chance. In most cases, a p-value of 0.05 (5%) is accepted to mean the data is valid.

2.2 Results

I performed T-test between group “new users with early win” and group “new users with early lose” based on D7 retention calculated in the manner below (**Table1**). The boxplot of Table1 is shown as **Fig.4**.

**Table.1** D7 retention in group “users with early win” and group “users with early lose"

| | win_new_D7 | lose_new_D7 | start_day | end_day |
|---|---|---|---|---|
| **0** | 0.086420 | 0.048991 | 2017-02-14 | 2017-02-20 |
| **1** | 0.078067 | 0.046595 | 2017-02-15 | 2017-02-21 |
| **2** | 0.095406 | 0.043896 | 2017-02-16 | 2017-02-22 |
| **3** | 0.099338 | 0.058997 | 2017-02-17 | 2017-02-23 |
| **4** | 0.118812 | 0.065534 | 2017-02-18 | 2017-02-24 |
| **5** | 0.106061 | 0.072165 | 2017-02-19 | 2017-02-25 |
| **6** | 0.109290 | 0.059226 | 2017-02-20 | 2017-02-26 |
| **7** | 0.073826 | 0.034682 | 2017-02-21 | 2017-02-27 |
| **8** | 0.073446 | 0.038760 | 2017-02-22 | 2017-02-28 |

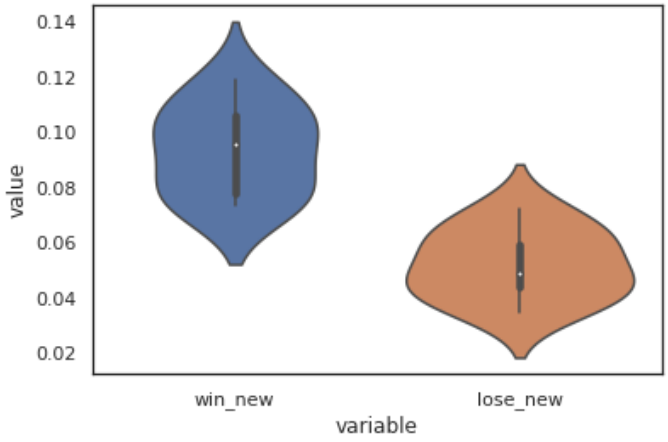


**Fig4.** Box plot on D7 retention in group "new user with early win" and "new user with early lose"

The t-test result gives us t statistics 5.983001182578937 and p-value 1.9127239678941084e-05, thus, we reject the hypothesis that group “users with early win” and group “users with early lose" have equal mean D7 retention, indicates that new players with better early performance have a better classic D7 retention.