A New Growth Model for Global Tech Startups

By Mario Morales (mario.morales@socialatomgroup.com), Prometheus Fellowship, Firstrock Capital VC

Winter 2017-2018

ABSTRACT

Growth and retention are central concepts in modern startup execution and valuation (see **(GRAHAM 2012)**, and **(HARIHARAN 2017)**). This paper uses concepts developed by venture capital firms “VC”, practitioners and professional investors to formalize and evaluate growth and retention observed in startups to develop functions and algorithms that identify features correlated to the main core (critical) action, optimal frequency of cohort analysis, top users and best-performing cohorts as a contribution to the literature and the practitioners.

Keywords: Startup, Cohort Analysis, Cohort Data Format, Firms Growth

JEL Classification: L26, M13, C81, C43

INTRODUCTION

This research focuses on modeling customer growth and retention for global startups and technology-based companies. The process of growth and the sources of differences in performance across startups are among the most important challenges in product development, venture capital investment decision making and technological public policy.

The primary purpose of this working paper is to formalize some concepts defined after the review of the state of the art in the literature on the subject (online sources, economic growth theories, mathematical developments on triangular matrices, cohort data analysis from social research, and longitudinal data analysis from statistics) and then provide a section of stylized facts (statistics) observed in the current literature.

The practical goal of our long-term research is the understanding of factors that will allow startups to improve their product and grow with the use of analytical tools designed for it.

Our final motivation extends beyond the methodological development and will seek to understand the fundamental causes of the theory and value of the firm in a disruptive digital economy (startup economy).

We have called our project **Lambdak**, and have defined certain algorithmic steps to execute our analysis.

In this study we:

- Propose a pipeline that allows startups to assess retention and growth based on a conceptual framework.
- Develop an algorithm that identifies the features that define the core (critical) action of the startup's application or system.
- Develop an algorithm that identifies the optimal frequency of cohort analysis for the target startup.
- Provide estimations of top users and best cohorts.
- Provide a data wrangling script to optimize the production of cohort datasets obtained from MIXPANEL.

CONCEPTUAL FRAMEWORK

According to our investment philosophy (inspired by ideas including **(GRAHAM 2012)**, **(HARIHARAN 2017)**, and **(A. SHIU 2017)**), in a digital economy the value of a company is measured by growth in a leading indicator KPI. The sustainable growth of this KPI then indicates positive demand for and engagement with the startup's product (Product Market Fit).

1. Leading Indicator KPI is defined by founders based on the market necessity, or problem being solved.

The first task is identifying the features that are correlated to our **leading indicator "KPI"**[1]. In this context, we take the KPI for growth as defined by the startup main goal, i.e. customer (user) count per period.

2. Feature Identification

- Once the KPI is defined, We build a vector of features based on the activity observed in the system and an index of order in a sequence of actions of execution.
- After this, we build a correlation matrix of these features and eliminate redundant ones with a collinearity measure.

3. Growth Function Construction (Growth Accounting)

As previously stated, our goal is to optimize a growth function for a defined KPI. While we reviewed the theory of growth models for startups (virality, growth accounting) established by practitioners over time from 2010 (see **(SKOK 2010)**, **(VOHRA 2012b)**, **(VOHRA 2012a)**, **(VOHRA 2013)**, **(CHEN 2014)**, **(QUINT 2014)**, **(HSU 2015d)**, **(HSU 2015e)**, **(ZORTIKIS 2016)**, **(SALOI 2016)**, **(GILL 2017)**, and **(A. SHIU 2017)**) we identify that some concepts of growth match frameworks developed historically in the economic theory (see for historical context and details: **(SOLOW 1956)**, **(SWAN 1956)**, **(ROMER 1986)**, **(ACEMOGLU 2008)**, and the online resources of the class about economic growth **(ACEMOGLU 2017)**)

In a basic framework, the change in the growth function of a defined KPI (i.e. user count, revenue per user) is explained by a change in the count of users that execute the core action feature that we found and is autoregressive, represented herein as:

$$C_t = \Delta C_{t-1}$$ Ct=ΔCt−1

$$where\ C_t{:}number\ of\ customers\ in\ the\ period\ t$$ where Ct:number of customers in the period t

.

In our framework, we can reconcile the theories observed as a Cobb Douglass Function defined by:

$$C_t = C_{t-1}^{\alpha_1} N_{t-1}^{\alpha_2}$$ Ct=Ct−1α1Nt−1α2

$$where\ C_{t-1}{:}number\ of\ customers\ in\ the\ period\ (t-1)$$ where Ct−1:number of customers in the period (t−1)

$$N_{t-1}{:}number\ of\ new\ customers\ in\ the\ period\ (t-1)$$ Nt−1:number of new customers in the period (t−1)

$$\alpha_1{:}\ is\ the\ observed\ coefficient\ of\ retention(churn)$$ α1: is the observed coefficient of retention(churn)

$$\alpha_2{:}\sum_1^n (invites * conversion\ rate)$$ α2:∑1n(invites∗conversion rate)

$$\alpha_2: \textit{is the observed coefficient of customer acquisition (virality) for n channels}$$
α2: is the observed coefficient of customer acquisition (virality) for n channels

While we observe in previous theories (see **(SKOK 2010)**, **(VOHRA 2012b)**, **(VOHRA 2012a)**, **(VOHRA 2013)**, **(CHEN 2014)**, **(QUINT 2014)**, and **(NACHUM 2015)**) the use of a deterministic function to model customer acquisition (virality) and will develop in posterior research theory and applied methods for it, for now we propose the use the empirical rates observed.

Due to the supermodularity properties of the Cobb Douglass Function (see **(TOPKIS 2011)**) we can use a log transformation to produce a linear function:

$$ln(C_t)=\alpha_1 * ln(C_{t-1})+\alpha_2 * ln(N_{t-1})$$
ln(Ct)=α1∗ln(Ct−1)+α2∗ln(Nt−1)

which is equivalent to the growth accounting decomposition proposed by **(HSU 2015d)**.

$$\Delta(C)=\alpha_1 * \Delta(C)C+\alpha_2 * \Delta(N)N$$
Δ(C)=α1∗Δ(C)C+α2∗Δ(N)N

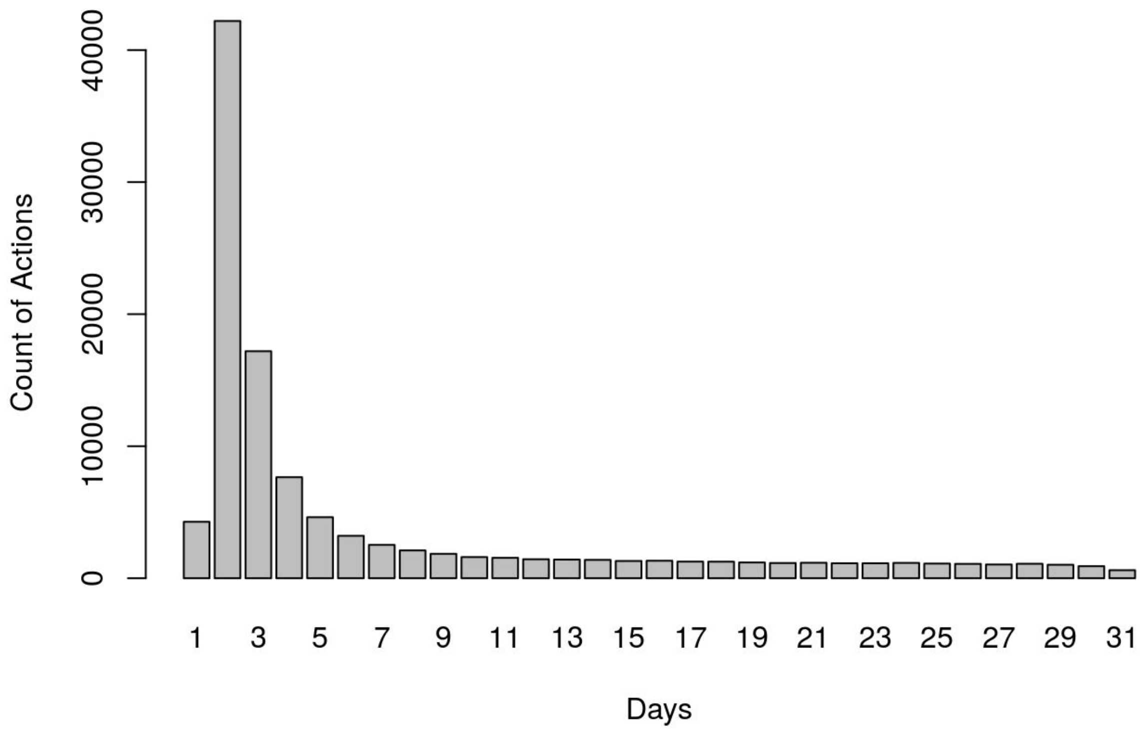
Within this context, we can observe this information using Cohort Methods. However, as we mentioned before we need to create theoretical models to estimate the rate of acquisition (virality) which is stochastic.

4. Identification of Optimal Frequency for Period of Analysis

At this moment we have identified the feature that we can filter to execute the counts in our analysis; however, we need to know the period of analysis or cycle of actions that characterize the use of our system. One possible way is to follow **(A. SHIU 2017)**. Algorithmically this means:

- Sort the dataset by customer (user) and the timestamp of the event.
- Compute the difference of time by user_id between the actions executed and express the value in minutes (days).
- Observe the histogram of the data and choose the period where all the data are clustered. (i.e In this graph below we see the actions mainly occur during a weekday)

FIGURE 1. Distribution of Frequency Core Action Executed over Time



5. Cohort Data Analysis

Cohort Analysis is a widely used methodological technique by venture capital firms (see **(HSU 2015d)**,**(HSU 2015e)**, **(HSU 2015b)**, **(HSU 2015c)**, **(HSU 2015a)**, **(HSU 2016)**, **(HARIHARAN 2017)**, **(MIXPANEL 2017)**, **(SHIU 2015)** and **(CARROLL 2017)**) to observe and study how often feature events occur in a given population of users. Feature events include click-throughs, page views, transacing, etc.

Cohort Matrices are triangular matrices that are composed in the rows by the cohort identification and then the columns show the periods over time. This structure is important because it allows us to introduce dynamical systems analysis from posterior research. (see **(ABARBANEL 1996)**)

cohort_period	2017-01-01	2017-02-01	2017-03-01	2017-04-01	2017-05-01	2017-06-01	2017-07-01	2017-08-01	2017-09-01	2017-10-01	2017-11-01
2017-01-01	121	109	104	94	93	84	84	78	76	72	70
2017-02-01		62	52	44	38	36	34	29	28	29	24
2017-03-01				38	35	32	34	27	21	22	19
2017-04-01				34	28	25	21	19	20	16	15
2017-05-01					35	25	26	25	24	17	15
2017-06-01						37	29	24	22	19	19
2017-07-01							45	37	34	32	29
2017-08-01								46	32	29	27
2017-09-01									45	37	30
2017-10-01										37	28
2017-11-01											22

Once the dataset is arranged in this form, a heatmap can be generated to show the intensity of the values in order to offer a clearer perspective of user / client behavior:

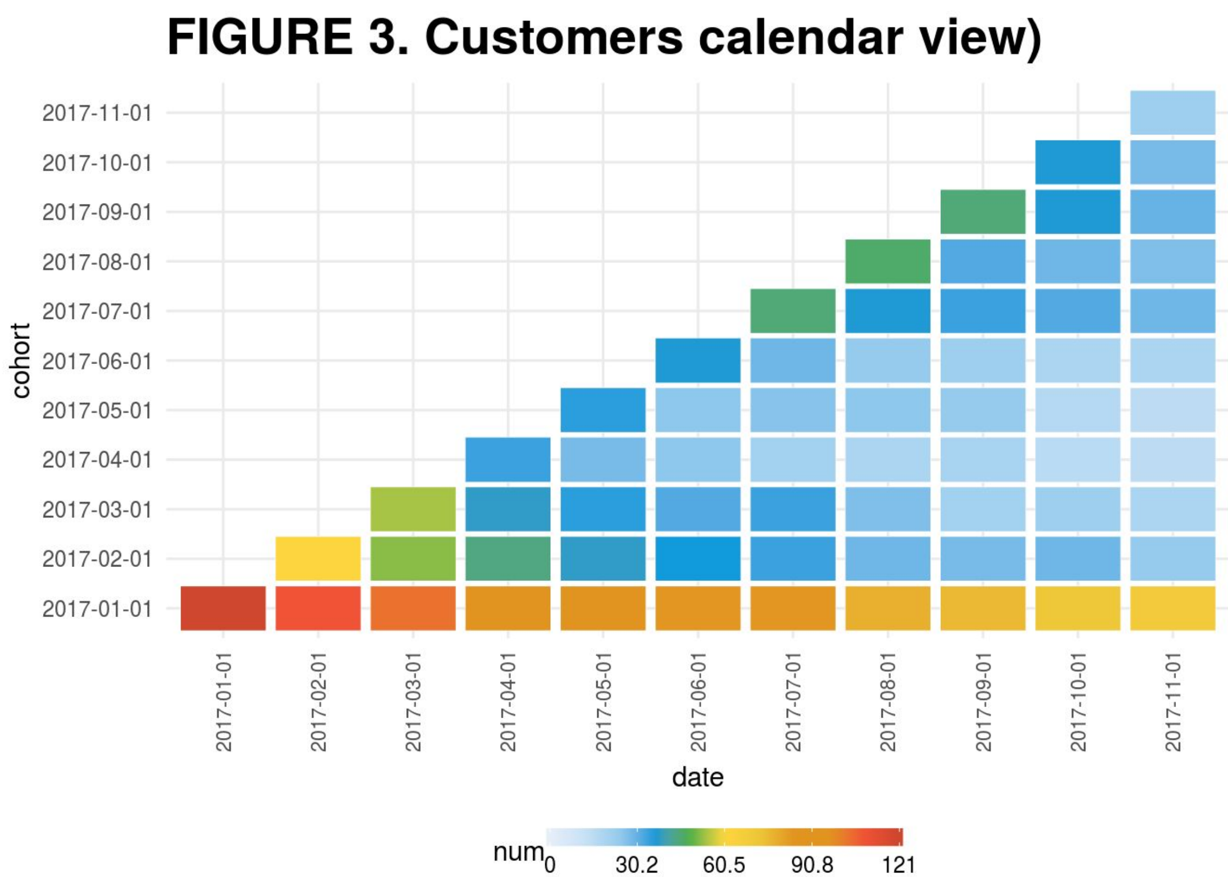
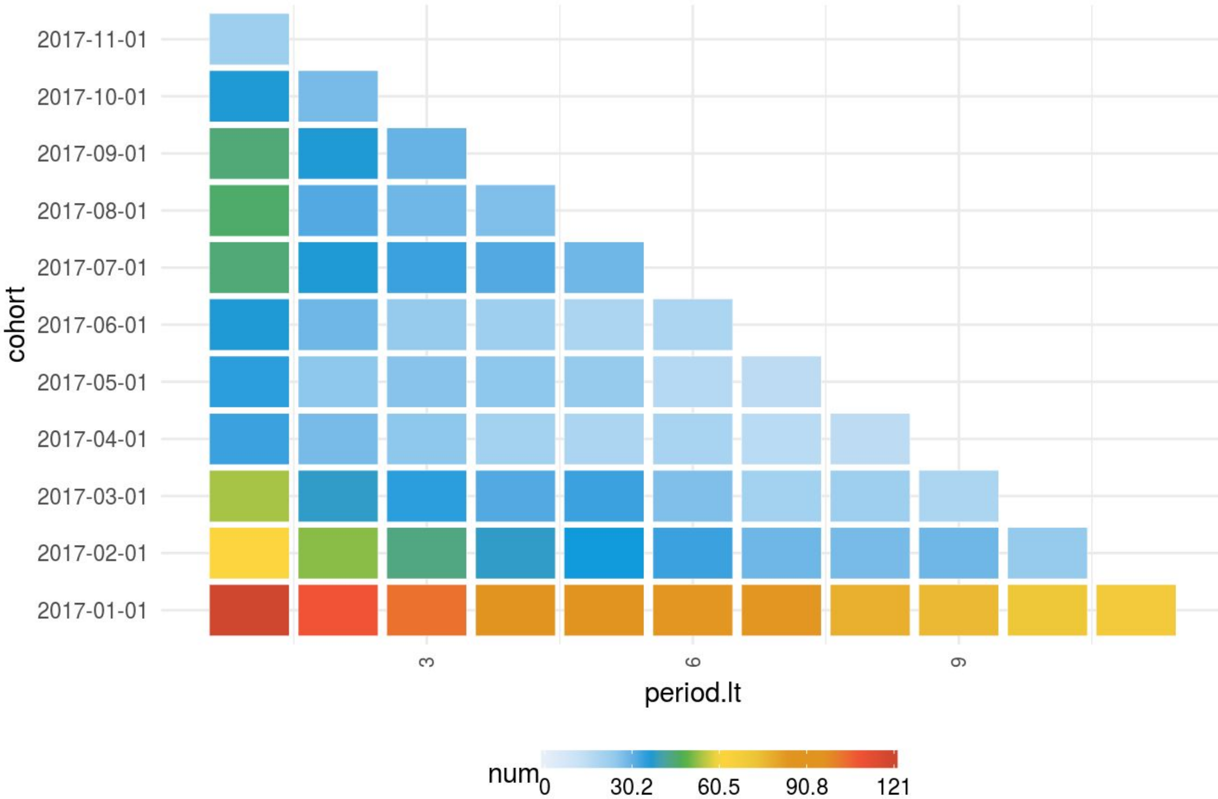


FIGURE 4. Customers lifetime view)



EMPIRICAL RETENTION OBSERVED IN THE LITERATURE

Certain benchmarks observed in the media and literature offer reference points for the retention expected of a successful startups by sector and business model:

FIGURE 5. BENCHMARK RETENTION PER VERTICAL SECTOR

Source : **(HARIHARAN 2017)**

Vertical	Period	Long Term Period	Long Term Target	Median
Social Network	Monthly	Month 12	45% – 65%	55%
On-Demand	Monthly	Month 12	20% – 30%	22%
Travel	Annual	Year 2	20% – 35%	29%
E-commerce	Monthly	Month 12	10% – 25%	16%
Subscription	Monthly	Month 12	25% – 35%	33%

FIGURE 6. BENCHMARK RETENTION 50 PERCENTILE GROWTH (TYPICAL)

Source : **(MIXPANEL 2017)** page 16

Vertical	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
FINANCE	34%	31%	21%	22%	18%	16%	17%	12%
SAAS	43%	33%	32%	30%	26%	22%	19%	17%
E-COMMERCE	17%	14%	11%	10%	8%	6%	5%	4%
MEDIA	33%	28%	24%	20%	18%	16%	14%	15%

FIGURE 7. BENCHMARK RETENTION 90 PERCENTILE GROWTH (ELITE)

Source : **(MIXPANEL 2017)** page 17

Vertical	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
FINANCE	61%	54%	48%	42%	35%	29%	27%	24%
SAAS	67%	60%	55%	50%	41%	42%	40%	35%
E-COMMERCE	72%	65%	58%	55%	48%	38%	34%	38%
MEDIA	56%	50%	44%	37%	32%	30%	28%	28%

DATA SOURCES

The traditional analysis the cohort behavior in the context of a technology-based startup is available from services including (“ADOBE” 1996),(“GOOGLE” 2005), (“FLURRY” 2005), (“PIWIK” 2007), (“MIXPANEL” 2009), (“AMPLITUDE” 2012),(“HEAP” 2013), (“CLEVERTAP” 2013), among others.

RECOMMENDATIONS FOR FUTURE RESEARCH

We are accumulating information to produce startup benchmark datasets per economic sector, activity, and startup business model. some exploratory data analysis in the existent benchmark datasets

APPENDIX 1: COHORT WRANGLING USING SQL EXECUTED AT INTERNAL DATABASES

For Internal datasets we suggest to use SQL, we suggest to check (PERISCOPE 2014), (CHARTIO 2017), (TAMURA 2016), (NGUYEN 2017), (HSU 2016)

APPENDIX 2: DATA EXTRACTION FROM MIXPANEL

(“MIXPANEL” 2009) is a current system that we use in this research for our analysis:

-When you log into MIXPANEL you will find a dashboard:

FIGURE 8

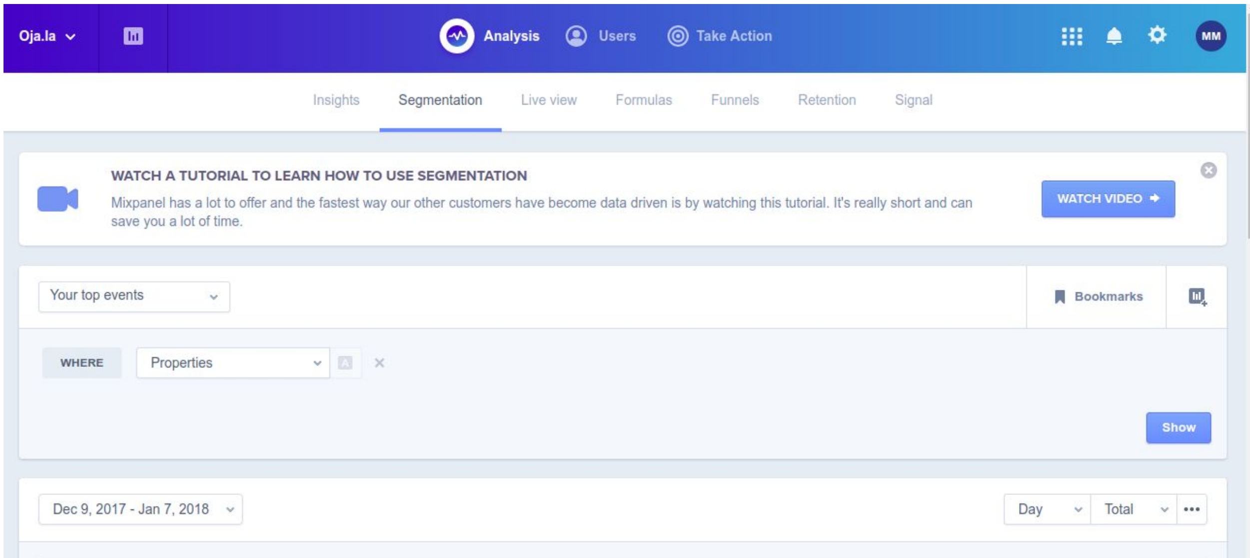


FIGURE 9 -Once you are there, you need to select the JQL option:

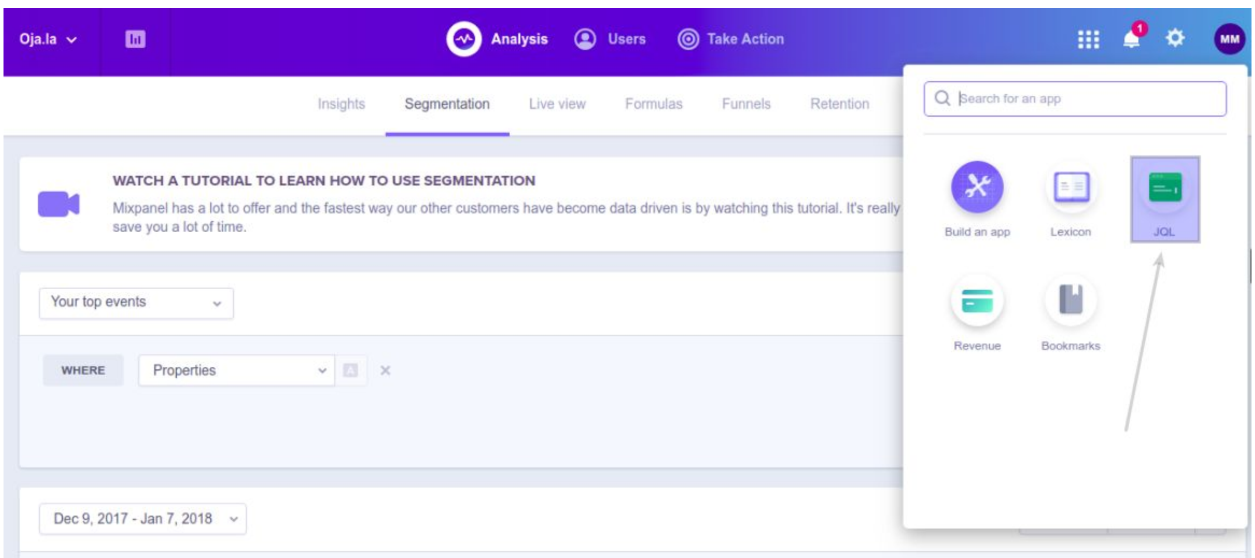
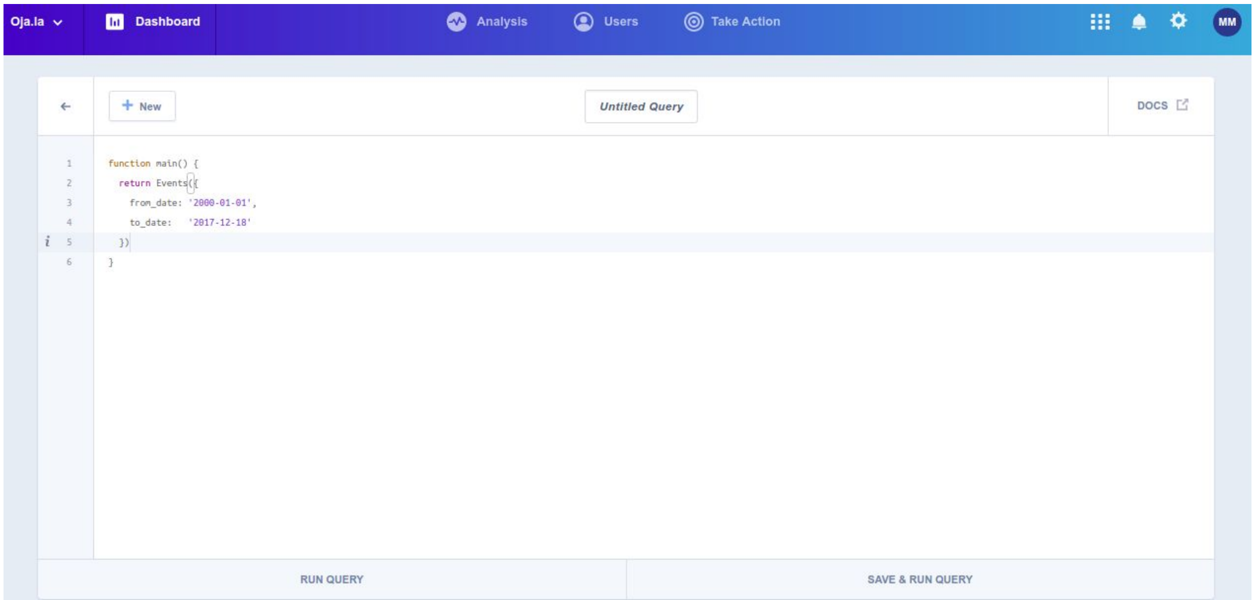


FIGURE 10 -The JQL script for data extraction is:



The data observed look like this:

date_confirmed	user_id	revenue
2017-01-24 13:53:37	10006	0
2017-01-25 0:14:37	10006	0
2017-01-25 0:34:35	10006	0
2017-01-25 0:39:10	10006	0
2017-01-25 0:40:06	10006	0
2017-01-25 0:40:53	10006	0

CONFLICT OF INTEREST

The author declares there is no conflict of interests regarding the publication of this paper.

ACKNOWLEDGMENT

This work is supported by the Prometheus fellowship (Winter 2017-2018). Prometheus is a three-month program tailored to engineers and applied data scientists that allows them to work and research in challenges observed in the startup environment which ends with contributions to global startup ecosystems and the creation of new companies and ideas. Prometheus is funded by Firstrock Capital.

REFERENCES

A. SHIU, AND A. MADHAVAN. 2017. "PRODUCT Analytics Playbook: Mastering Retention: Proven Methods for Turning User Behavior Insights into Retention Strategies." november. https://www.productanalyticsplaybook.com/.

ABARBANEL, H.D.I. 1996. *Analysis of Observed Chaotic Data*. Springer-Verlag: Institute for Nonlinear Science.

ACEMOGLU, D. 2008. *Introduction to Modern Economic Growth*. Princeton University Press.

———. 2017. "14.452 - Introduction to Economic Growth." https://economics.mit.edu/faculty/acemoglu/courses.

"ADOBE." 1996. http://www.adobe.com/data-analytics-cloud/analytics.html.

"AMPLITUDE." 2012. https://www.amplitude.com/.

CARROLL, A. 2017. "Capital as a Service a New Operating System for Early Stage Investing 2017." october.https://medium.com/social-capital/capital-as-a-service-a-new-operating-system-for-early-stage-investing-6d001416c0df.

CHARTIO. 2017. "Performing Cohort Analysis Using Mysql." https://chartio.com/resources/tutorials/performing-cohort-analysis-using-mysql/.

CHEN, A. 2014. "Retention Is King." june. http://andrewchen.co/retention-is-king/.

"CLEVERTAP." 2013. https://clevertap.com/.

"FLURRY." 2005. https://developer.yahoo.com/analytics/.

GILL, W. 2017. "How to Build a Growth Model." march. https://mobilegrowthstack.com/how-to-build-a-growth-model-a7dd9d9a68a7.

"GOOGLE." 2005. https://www.google.com/analytics/.

GRAHAM, P. 2012. "Startup=Growth." september. http://www.paulgraham.com/growth.html.

HARIHARAN, A. 2017. "Growth Guide: How to Set up, Staff and Scale a Growth Program." july.http://blog.ycombinator.com/growth-guide2017/#checkretention.

"HEAP." 2013. https://heapanalytics.com/.

HSU, J. 2015a. "Diligence at Social Capital Part 5: Depth of Engagement and Quality of Revenue." november.https://medium.com/swlh/diligence-at-social-capital-part-5-depth-of-usage-and-quality-of-revenue-b4dd96b47ca6.

———. 2015b. "Diligence at Social Capital Part 3: Cohorts and (Revenue) Ltv." october.https://medium.com/swlh/diligence-at-social-capital-part-3-cohorts-and-revenue-ltv-ab65a07464e1.

———. 2015c. "Diligence at Social Capital Part 4: Cohorts and (Engagement) Ltv." october.https://medium.com/swlh/diligence-at-social-capital-part-4-cohorts-and-engagement-ltv-80b4fa7f8e41.

———. 2015d. "Diligence at Social Capital Part 1: Accounting for User Growth." september.https://medium.com/swlh/diligence-at-social-capital-part-1-accounting-for-user-growth-4a8a449fddfc.

———. 2015e. "Diligence at Social Capital Part 2: Accounting for Revenue Growth." september.https://medium.com/swlh/diligence-at-social-capital-part-2-accounting-for-revenue-growth-551fa07dd972.

———. 2016. "Diligence at Social Capital, Epilogue: Introducing the 8-Ball and 'Gaap for Startups'." february.https://medium.com/swlh/diligence-at-social-capital-epilogue-introducing-the-8-ball-and-gaap-for-startups-7ab215c378bc.

"MIXPANEL." 2009. https://mixpanel.com.

MIXPANEL. 2017. "The Product Benchmarks Report 2017." october. https://mixpanel.com/2017-product-benchmarks-report/.

NACHUM, G. 2015. “Dissecting Virality — the Mathematical Formula.” october. https://readwrite.com/2015/10/09/virality-math-formula/.

NGUYEN, H. 2017. “Quick Guide: Calculate Cohort Retention Analysis with Sql.” october.https://blog.holistics.io/calculate-cohort-retention-analysis-with-sql/.

PERISCOPE. 2014. “How to Calculate Cohort Retention in Sql.” june. https://www.periscopedata.com/blog/how-to-calculate-cohort-retention-in-sql.

“PIWIK.” 2007. https://piwik.org/.

QUINT. 2014. “Quint Growth Retention Model.” june. http://quintgrowth.com/growthmodel.html.

ROMER, P. 1986. “Increasing Returns and Long-Run Growth.” *Journal of Political Economy* 94 (5): 1002–37.https://www.jstor.org/stable/1833190.

SALOI, M.J. 2016. “A Simple Growth Model for Your Start-up.” april. https://hackernoon.com/a-simple-growth-model-for-your-start-up-de5e9315541.

SHIU, A. 2015. “Why You Need Cohorts to Improve Your Retention.” november. https://de.com/blog/2015/11/24/cohorts-to-improve-your-retention/.

SKOK, W. 2010. “Lessons Learned – Viral Marketing.” march. http://davidskok.ulitzer.com/node/1310273.

SOLOW, R.M. 1956. “A Contribution to the Theory of Economic Growth.” *Quarterly Journal of Economics* 70 (1): 65–94.https://academic.oup.com/qje/article-abstract/70/1/65/1903777.

SWAN, T.W. 1956. “Economic Growth and Capital Accumulation.” *Economic Record* 32 (2): 334–61.http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4932.1956.tb00434.x/abstract.

TAMURA, K. 2016. “Rolling Retention Done Right in Sql.” july. https://blog.treasuredata.com/blog/2016/07/22/rolling-retention-done-right-in-sql/.

TOPKIS, D.M. 2011. *Supermodularity and Complementarity*. Princeton University Press.

VOHRA, R. 2012a. “How to Model Viral Growth: Simple Loss.” october. https://www.linkedin.com/pulse/20121002124206-18876785-how-to-model-viral-growth-the-hybrid-model/.

———. 2012b. “How to Model Viral Growth: The Hybrid Model.” october.https://www.linkedin.com/pulse/20121002124206-18876785-how-to-model-viral-growth-the-hybrid-model/.

———. 2013. “How to Model Viral Growth: Retention & Virality Curves.” april.https://www.linkedin.com/pulse/20130402154324-18876785-how-to-model-viral-growth-retention-virality-curves/.

ZORTIKIS, M.J. 2016. “How We Replicated Amazon’s Growth Model for Our Startup: The Growth Pitfall.” May.https://medium.com/inside-viral-loops/how-we-replicated-amazons-growth-model-for-our-startup-b7ab89252ee9.

---

1. (**Key_Performance_Indicator “KPI”** : Is a quantifiable variable that measures the performance over time of the capability of a company to achieve a strategic or operational goal and reflect the financial value of a company and can be used to compare it against similar companies)
2. Finally, we execute a correlation analysis with the target KPI and assess the top X% positive and X% negative values to identify features that affect the growth of the target KPI. (A proposal to make this stage automatic is the use of principal component analysis, however for current interpretation we believe the decision maker can determine the most effective action.)