Contents

1. Time Line

- Pre-1980's: The Execution Era
- 1980-2000: Porter's Forces
- 2000-2018: Information Rules
  Therefore data is oil (Data Moats)
- 2018-2020: Algorithmic Moats
- 2020 - Present: Algorithms are Consulting Firms
- 2022: What now...

2. Nick's Take

- There are LOTS of possible data moats (be algorithmic, special data, etc.), but
  (1) You have to get in very early to get a huge head start to be able to realize it.
    - Consider Netflix, Hulu, Youtube, Paramont Plus, HBOMax
    - Netflix had a HUGE head start before other companies got involved. That head start afforded them significant protection.
  (2) These moats tend not to last as long as expected due to the issues we have seen before:
    - Marginal value of additional data
    - Marginal cost of additional data
    - Data staleness
    - Rise of pretrained models (GPT-3, Huggingface, etc.)

   (3) Proprietary data products (lexis-nexus, ZoomInfo, Dun and Bradstreet) are an additional type of moat, protected by "owning" the data.

      These moats are very different than the moats we spoke of before and usually require an additional aspect (protection by government, deep IP, a huge inventory of costly to create data).

- In essence, I'm pretty down on "data moats" of any type. Like the a16z article said – data isn't magical and shouldn't be relied upon to create a sustainable business.
- I am very bullish on data as an accessory or accelerator of other moats.
- Given this, what types of "data"/"ai" companies are likely to be successful **and be good for data scientists to work at?**
  (1) Those with an obvious value proposition which requires data.
    - Ambient.ai (security vision company). Note that I'm higher up on "profit center" data than "cost center" data (dunnhumby / loyalty programs)
  (2) Companies whose value proposition is outside a company's core competency
    - Threat of integration lower in these situations
    - Assert.ai (log analysis). Security / Uptime.
  (3) Data support propositions: Companies which directly support data operations
    - Companies which provide direct support to data operations, thereby off loading work from core teams.
    - Companies that do logging, ETL, data storage, security, data version control and model version control.
- What about AI Research?
  - I'm lukewarm. Companies like open.ai, hugging face, etc.
  - Cost of building these models is incredibly high and as technology and techniques change quicker and quicker they do not have as much of a head start as may first appear.

3. Unit Economics

3.1. **Sam Altman Article.**

- Who is Sam Altman?
- Former President of YC, current CEO of OpenAI
- Well known Venture Capitalist
- Article is from 2015, but speaks to the importance of "Unit Economics"
- What is Unit Economics?
  - It is how we describe the basic profitability on a "unit sold" basis.
  - Gets to the heart of "profitability"
  - We need to think about unit economics, especially in start-up world because:

(1) A lot of companies lose money and it provides a framework to understand if there is a way out of this.

(2) There is (very frequently) a mismatch between the timing of revenue and the timing of costs associated with that revenue.

- To explore these we need to have a basic "model of the firm"

3.2. Model of a firm's profit #1.

- Consider the (simple) case of a manufacturing firm which sells widgets.
- For a firm like this we, at the simplest level define the firms profitability using the formula: $\Pi = P \cdot Q - V \cdot Q - F$
  (1) Π Profit
  (2) P Price : What is charged for the unit
  (3) Q Quantity Sold: How many are sold
  (4) V Variable Costs: Costs which change with the number of units sold.
  Examples (Sales Team, raw materials)
  (5) F Fixed costs: Costs which are static and independent of the number of units sold.
  Examples (HR, Rent)
- When we talk about unit economics we are often talking about the **Margin** which is defined either as a percent or as a nominal value:

$$\frac{P-V}{P}$$
$$P-V$$

- Assuming that your margin is positive then there is a level of production which will be break-even and, any quantity above that will yield positive profit:

$$\Pi = 0 = PQ - VQ - F$$
$$F = Q(P-V)$$
$$Q^{BE} = \frac{F}{P-V}$$

- This is called the "CVP" Model
- This model makes a ton of sense for many companies, but not all.
- For startups where the CVP model makes sense, the notion of "unit economics" is trying to predict how P, V and C are going to change as a product grows and changes.
- We have to project, based on what we see in the company's financials to understand if there is the possibility of profitability. Only a few things we can do:
  - Are there efficiencies that will lower V?
  - Can we charge more (raise P) as our product gets better?

  - Can we expand to additional markets (increase Q)
  - Can we change our fixed costs (lower F) – tends to be the least likely.

3.3. **Unit Economics for Start-ups (Model #2).**

- For a traditional firm with simple transactions this is a pretty straightforward calculation.
- There are some features that tend to blow this up for startups:
  (1) Startups make "growth" decisions that may negatively impact short term profit.
    - Hiring roles when trying to find product market fit
    - Investing in "potentials" that may or may return dividends

  (2) Q maybe too low because trying to build.
    - Because the company is just starting out we don't expect them to be profitable.
    - Initial units are often far more costly until infrastructure put in place, economies of scale hit.

  (3) Repeated Revenue / Subscription Revenue (LTV) and Costs (CAC) for information goods.
    - Information goods – tend to have near zero marginal costs, BUT have reoccurring revenue
    - Examples: advertising for facebook, Netflix, google ads
    - In this situation we need a different model for our company since the CVP model of the company isn't necessarily well defined and will not represent the underlying economic conditions.
- On that last point, consider the case where a user purchases a 6 month subscription to a digital good. So each month they pay 50 cents per month. We run an advertising campaign which gets us new users at $1 a piece (100 of them). If we look at the P&L for this we would see Table #2 Below.
- What if we decide to purchase an additional 100 users in month #2? We would see Table #3
- In Month #2 we actually look worse in the 2nd case!
- It's pretty easy to see that by manipulating advertising spend we can push around Profits and Losses pretty easily.
- Because of this the CVP model isn't as important for many technology and data companies.

3.4. **LTV/CAC Model.**

- The underlying reason that the CVP model breaks down for information/subscription goods is that the appropriate "Q" isn't number of units, but is now a "customer".

Table 1. Only purchase first month

Month	revenue	cost	New Users	profit
1	100*.5	100*1	100	-50
2	50			50
3	50			50
4	50			50
5	50			50
6	50			50

Table 2. Purchase more users

Month	revenue	cost	New Users	profit
1	100*.5	100*1	100	-50
2	50	100*1	100	0
3	100			100
4	100			100
5	100			100
6	100			100
7	50			50

- Think about netflix – "Q" is a customer, not a movie.
- We want our unit economics to not be calculate on a unit sold, but on a customer served. For traditional companies these maybe the same, but for tech companies they are often different.
- How do we calculate Unit Economics in this case?

$$\text{Margin} = \text{LTV} - \text{CAC} - textCOGS$$

- LTV : **Life-time value** of a customer – The total value that a customer generates
- CAC: **Customer Acquisition Costs** – the cost of getting that customer (advertisement, sales, etc.)
- COGS: **Cost of Goods Sold** – If there is a physical deliverable this is what is included (e.g. the meals of blue apron)

3.5. Calculating LTV.

- (1) "Monetization"/"Revenue curves and (2)"Retention"/"Churn" Curves in order to estimate long term customer profitability.
- In general we create two curves:
  (1) Retention Curve
    - This represents the probability that a user returns in a specific time interval.
    - Tons of ways of calculating it, but usually at a single point, so the formula would look something like this:

$$R^{30}_{1-1-2020} = \frac{\text{\# of People returning on 1-30-2020}}{\text{\# of People who installed / signed up on 1-1-2020}}$$

    - This is the Day 30 Retention for users who signed up on 1/1/2020.

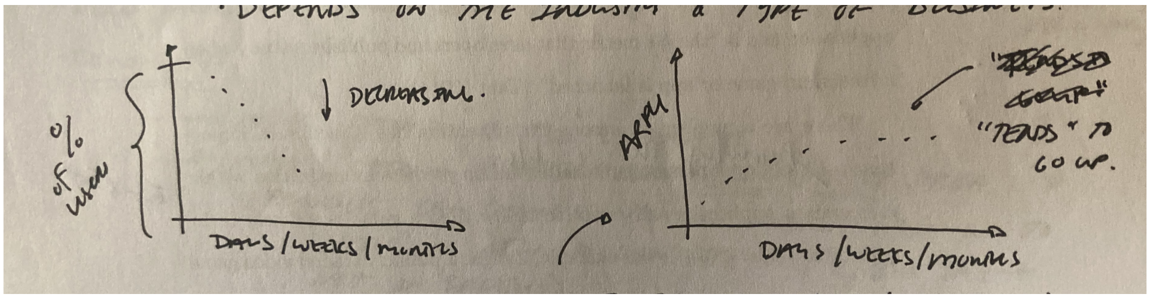
  (2) Revenue (ARPU) Curve
    - This represents the amount of money an average user generates at a particular point in their customer journey.

– Tons of ways of calculating it, but usually at a single point, so the formula looks something like this:

$$ARPDAU^{30}_{1-1-2020} = \frac{\text{Amount of money generated from users who installed on 1-1-2020 on 1-30-2020}}{\text{Total Users who used the application / were still using the application on 1-30-2020}}$$

– In Freemium world we call the denominator the "DAU" or Daily Active Users. If we were doing this at a monthly interview we would call the denominator the "MAU" or Monthly Active Users.

(3) We can build curves from the above by varying the days since install.



- If we multiple these curves together we get the total revenue generated from a particular *cohort*
- A cohort of users is a set of users who installed at the same time and have other essential properties which are known at the time of sign up or install (geographic location, device used).
- We (essentially) multiple these two curves together to get the total revenue generated per user of that cohort.

TABLE 3. Retention by Day

Day	Retention
0	100%
1	50%
2	30%
3	10%
4	7%
5	.
6	.

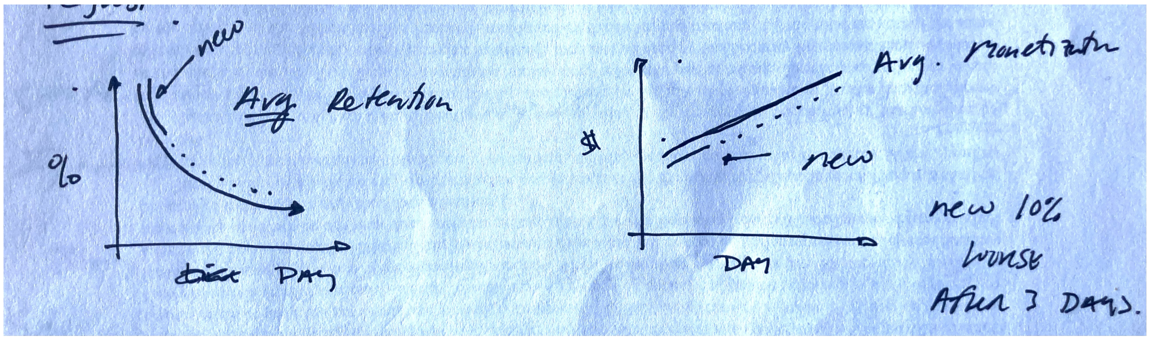
TABLE 4. Revenue Per Active User Per Day

Day	Revenue
0	$.30
1	$.32
2	$.33
3	$.35
4	.
5	.

$$LTV = \sum_{i=0}^{n} \text{ARPDAU} \cdot \text{Retention}$$
$$= .3 + .32 * .5 + .33 * .3 + .35 * .1$$

3.6. Predicting LTV.

- When trying to estimate or predict LTV we take whatever information we have and try to adjust the curves above to match what we believe. This is difficult because it is forward looking information.
- For example, if we have a new cohort of users and their retention has been 20% higher than the "model" for the first three days while their revenue per user is 10% worse than our model.



- We almost always want to build this based on groups of similar users. Commonly used "segments" include:
  - Geography
  - Language
  - Device
  - Acquisition source (Facebook Ad? Google SEO?)
- One of the hard parts of building predictive LTV models is balancing the need for specificity on cohort against the available sample sizes. In my experience there is usually not enough volume to break down by every desired dimension, so you scale up cohorts accordingly.
- For example, you may only have a few users from Brazil, so you may want to group them with users from Portugal (language) or for users from South America (geography).

3.7. Model Summary.

- We have two different models of unit economics, one based on the unit being a consumer (LTV) and one based on a unit being something sold (CVP).
- Both are useful for understanding the profitability of a start-up on the unit economic level.
- Depending on the business itself (usually) only one makes sense.
- Returning to Sam Altman's Article:

- Doesn't really specify which model, but what he says applies to both (I actually find this in a lot of popular business writing):
  If $P - V \leq 0$ then the company will never be profitable without changes.
  If LTV – CAC – COGS ≤ 0 then then company will never be profitable without changes.
- Long term, companies have to get these under control in order to be profitable.

4. Margins Matter Article

- Written by Tomasz Tunguz
- VC at Redpoint Ventures
- Pretty well known (and free!) newsletter
- Article looks at a company's **unit economics** with the CVP model and shows that small changes in the underlying parameters (P and V, specifically) have a huge impact on a company's runway.
- Important take-away is that companies and VC still continue to use the CVP model, it's not all LTV and CAC.

5. Blue Apron Articles

- Two articles, both written by Daniel McCarthy
- Professor of Marketing at Emory University
- Looks at the Blue Apron S-1 (what is an S-1?)
- Importantly, financial docs are more inline with the CVP style framework.
  Revenue is recognized at the time the service is provided
- What do these articles do?
  They back out the underlying LTV/CAC from the CVP data
- Specifically, Daniel built a model which allows his to estimate the parameters of the CVP Model (Monetization, Retention and CAC) and then understand the unit economics from this model's perspective.
- Conclusions:
  (1) Retention is bad – 72% of customers gone after six months
  (2) Revenue of new customers < revenue of old customers
  (3) Revenue curves are going *down*
  (4) Customer Acquisition costs are increasing
- I read the underlying papers for how he creates the model to move from financial numbers to LTV based numbers and while he has to make a number of assumptions they are relatively reasonable. I don't think that his conclusions are drawn from a bad model.

- Extensions (if time): Porter Analysis of Blue Apron.