

Davis Balaba

Data-Driven Business Growth

Course Notes

365[✓]DataScience

Table of Contents

1. Introduction.....	4
2. The stages of data maturity.....	5
2.1 The stages of data maturity and what you will see next.....	5
2.2 How to go from no data to some data	5
3. Data maturity level 1	6
3.1 Intro to project 1.....	6
3.2 Why do the analysis?	7
3.3 Formulating an analysis plan	7
3.4 The data we will use.....	8
3.5 Exploring the data: large dataset.....	8
3.6 Exploring the data: small dataset.....	8
3.7 Customer journey.....	9
3.8 Top of funnel opportunities.....	10
3.9 Middle of funnel opportunities	10
3.10 Lower funnel opportunities	11
3.11 Test and learn	13
3.12 How to get to data maturity level 2	13
3.13 How to ask for more funding.....	14
4. Data maturity level 2	15
4.1 Intro to Project 2.....	18
4.2 Crawl stage	18
4.3 Walk stage.....	19
4.4 Run stage.....	21

4.5 What is an A/B test.....	21
4.6 Statistical power	22
4.7 Sample size	22
4.8 Test power vs lift.....	22
4.9 How to get to the next level.....	23
4.10 How to ask for funding from decision makers	24
5. Data maturity level 3	25
6. What to do to improve after level 3	27

Abstract

What is the growth mindset? Why was it such a revolution?

The growth mindset manifests itself in a culture of discontent of the current state.

The underlying assumption is that there is always more value to be uncovered and data is the path to unlock that value.

What if our business does not have as much data as a large tech company?

1. Growth is a mindset. Data fuels your growth thinking
2. You don't realize how much data you have until you start focusing on it

The course provides actionable advice on how to get started on your data journey.

You don't have to have big data to be data-driven.

Why is the growth mindset important?

1. Ensures that you always work on the most important problems first
2. Helps clarify what outcomes you can expect
3. Focused on the desired end-result and encourages an execution mindset

1. Introduction

A growth mindset is a way of thinking, which requires 3 steps:

I. Devise clear hypotheses

(Example: "If investing in influencer-based marketing is beneficial to the top line then we will see higher than average ROI compared to other forms of marketing")

A hypothesis has a clear:

- a) Value statement
- b) Measurable outcome
- c) Rationale for why

II. Explore and rank hypotheses

Criteria to evaluate and determine which hypothesis to pursue:

- Competition
- Expertise
- Ease of adoption

III. Testing

Test fast and learn faster

Having set up a repeatable process of going from hypotheses to learnings, one can start to see how much higher incremental impact the growth mindset has compared to less rigorous and disciplined approaches.

Why the growth mindset is important?

- a) It ensures you always work on the most important problems first
- b) It helps clarify what outcomes you can expect
- c) Focused on the desired end-result and encourages an execution mindset

Does the growth mindset apply to all industries and verticals?

Yes, it does. Any vertical in which you can have multiple challenges and possible solutions can benefit from the growth mindset.

2. The stages of data maturity

2.1 The stages of data maturity and what you will see next

Data maturity is defined by how much data you have and your ability to access and take advantage of it.

Data maturity stages:

- Stage 1 (beginner) - no data or limited and sometimes untrusted data.
- Stage 2 (intermediate) - small scale data; you can perform analyses; report findings; perform some optimizations on existing projects.
- Stage 3 (advanced) - data at large scale compared to those in your vertical; End-to-end functionality to manage the data and innovate from your data-driven findings.

The higher your data maturity level, the more hypotheses you can test. The more impact you can have.

2.2 How to go from no data to some data

When you have no data, there are a variety of techniques that can be used to get insights. It all boils down to asking your customers in some way, shape, or form.

3. Data maturity level 1

At this stage you have some data but are not extracting much value from it.

For example:

You have data surrounding what people click on your website, but you are not using the data to make any decisions about what products to focus on or who to market specific products to in a systematic way.

Actions taken at this stage can yield remarkable results because you have not done many optimizations.

Often times the data offers clues what to focus on. You can use them to stack rank different projects based on factors such as:

- Opportunity size
- Cost
- Risk

Only by understanding your current state of affairs can objectively determine if the effort and resources you put into activities have a good return on investment.

Keep in mind that when you have a few data points there is a tendency to skew and not use truly representative data. So, this is important to consider when making decisions.

3.1 Intro to project 1

Company: Puppy Toys Inc

Description: Online retailer of puppy toys

Marketing strategy: Runs marketing campaigns on online platforms to drive potential customers to its website

Hypothesis:

We believe that by analysing the customer journey we can identify the weak points in our funnel.

Focussing our actions on these weak points will drive incremental sales.

Customer journey:

1. See ad
2. Click ad
3. Visit site
4. Click buy now
5. Add to cart
6. Purchase

3.2 Why do the analysis?

Before starting the analysis, it is important to understand the reason why you do it.

This enables you to identify a clear outcome you are aiming for.

This is important because, at first, when you get a hold of data you can go in many different paths. But clearly only a few of them will lead you to an outcome, which is meaningful, actionable and beneficial to the business.

If you set certain benchmarks before starting the analysis, you can start to hypothesize what the biggest opportunities for impact might be.

3.3 Formulating an analysis plan

Incremental sales – revenue that would not have happened if we had not taken certain actions

Analysis plan:

1. Understand the data
2. Understand how clean the data is
3. Summarize the data and generate benchmarks

When we understand benchmarks, we can compare to our own expectations and based on that we can start hypothesizing where the biggest opportunities are.

3.4 The data we will use

One caveat of spreadsheet tools is that they have a limitation of the number of rows that can be used.

3.5 Exploring the data: large dataset

A **data dictionary** defines every variable (every field) that you have in the data. It tells you what it means and what values it is supposed to take. This might be useful when starting your analysis, but it will be even more useful at a later stage when after some time you or someone else comes back to the analysis to refresh or edit it.

Data improvement plan - actionable steps that can be taken to improve data availability and quality in the future when more resources are available

3.6 Exploring the data: small dataset

Typically, you will use the insights generated with the smaller dataset to justify getting more investment to get more sophisticated tools and get more talented people on your team to accelerate your data journey.

3.7 Customer journey

Conversion rate: the percentage of targets that perform the action that you are trying to drive them to perform

Example:

An ad is shown to 100 customers. 5 of them click on the ad. Then the ad's conversion rate is $5/100 = 5\%$

Drop-off rate: number of customers who did not perform the intended action.
 $95/100 = 95\%$

Journey for customers who saw ads:

See ad	13.7k	79% of people who saw the ad didn't click
Click ad	2.86k	8% of people who clicked didn't visit the site
Visit site	2.63k	67% of site visitors didn't click 'buy now'
Click buy now	877	43% of 'buy now' clickers didn't add anything to the cart
Add to cart	497	31% of those who added to cart didn't purchase
Purchase	344	

When it comes to funnels, an X% improvement at any point of the funnel results in the same absolute impact at the bottom i.e. sales

The question to answer is: Where is the biggest opportunity for improvement, that is also realistically attainable?

3.8 Top of funnel opportunities

Investigate ad click-through rate and compare with competitors.

Beware of surgical targeting (having too narrow targeting that results in very high CTR). The downside of this strategy is your basket of opportunity is very narrow.

You might be better off diluting the CTR but expanding to a larger audience.

Conclusion: There is a huge gap upside in increasing the number of people who see our ads

Hypothesis: We believe that we can double our reach by advertising more via paid search and paid social. The incremental reach will result in incremental sales.

Risks:

CTR will decrease. This will increase cost per sale.

Current margin per unit will decrease. Despite this, overall profit might increase if we sell more units.

3.9 Middle of funnel opportunities

The question that needs to be answered is - "Is our site optimized to sell?"

Observations:

- The "Buy now" button is only located on 2 of 11 pages on our site
- It is rendered as a bright-green button that sits at the top right of the site and is always visible
- 40% of site visitors do not visit a page where we have "Buy now"
- Of the 2.63k site visitors, only 1.6k see the buy button
- Since 0.8k click, the true "Buy now" conversion rate is about 56% i.e. the drop-off rate is 44%

Conclusion: There is some potential upside in improving the accessibility of the “Buy now” button

Hypothesis: We believe that we can increase ‘buy now’ clicks by adding the button to every page. This will result in incremental “Buy now” clicks, which will lead to incremental sales

Risks: If we added a “Buy now” button to every page, we would expose an additional 1k customers to it. These customers are likely lower intent so we can assume a lower conversion rate for estimation. Let’s assume their “Buy now” click rate is 20% that of the higher intent users.

So 1k users with a “Buy now” click rate of 10% means we can have an additional 100 buy now clickers

Since 42% (344/825) of “Buy now” clickers make a purchase, we estimate an incremental 42 purchases.

3.10 Lower funnel opportunities

Click “Buy now”	877
-----------------	-----

Add to cart	497
-------------	-----

43% of “buy now” clickers did not add anything to the cart.

How can we reduce journey abandonment?

Users who click “Buy now” have shown some intent to buy. From our research cart abandonment rate for our vertical is around 30%, so we are underperforming. We need to figure out why users abandon the journey.

Possible reasons:

- CTAs are not compelling enough; Action: Test

- Friction in our product selection process; Action: Compare to competitors and identify best practices to test out
- Payment options are not trusted; Action: Compare to competitors and identify best practices to test out

Observations:

- Payment methods are on par with best-in-class practices
- We have a very low-pressure sales strategy
- Items are sometimes not preserved in the cart when the user navigates away
- 10% of users who have 'add to cart' end their session with an empty cart

Conclusion: There is an opportunity to improve the 'add to cart' process by re-engaging our site and trying alternative higher-pressure messaging.

Hypothesis: We believe we can improve our add to cart rate by introducing time-sensitive discounts. This will lead to incremental add to carts and more sales.

We believe that re-engineering our site to preserve items added to cart will reduce user frustration. This will lower cart abandonment and result in incremental sales.

Risks:

- Some users may just have a longer more deliberative purchase journey.
- Adding new elements to the site could lead to new bugs and breaks in the purchase flow.
- Effort vs Impact analysis
- Define the effort needed to execute an idea and the impact (in terms of increase in sales) it can have.

3.11 Test and learn

A/B testing is the gold standard

- Quantify true causality and thus incrementally
- Not always feasible

Other ways to approximate impact

- Pre/post analysis
- Time-series
- Diff-in-diff
- Causal inference

Clean A/B testing is not always possible

1. Evenly splitting target audience not always possible due to
 - a. All or nothing channels - TV
 - b. Channels where user-splitting is not possible
2. Randomization of test and control
 - a. Challenging if you can't identify users uniquely
 - b. Challenging if user identifiers can evolve over time or can be hidden from you
 - c. Challenging if you can't cleanly separate test groups from control groups so they don't influence each other

3.12 How to get to data maturity level 2

There is an exponential relationship between data maturity and impact. Putting in the effort to have more or higher quality data can unlock a disproportionate

amount of additional capabilities. It is no wonder tech companies are obsessed with data.

One of the first thing you can do when you get to data maturity level 2 is collect more data. A very important question here is 'What is the most valuable data to collect?' 'What data can you collect that will help you answer strategic questions?'

Example of missing data:

Missing steps in funnel of website events

Visits website

-missing step-

Goes to 'pricing' page

Checkout

Sometimes simple tweaks can get you to a much better place:

- Provide data in a different format
- Improving the quality of your data

Fallacy of complexity or Occam's razor: Denies the assumption that the more complex your solution the greater the impact.

3.13 How to ask for more funding

1. Demonstrate a proof of concept
2. Demonstrate the limitations of your current data infrastructure

Proof of concept:

- Why does your role or team exist and how does your role contribute to key company goals?
- What are the different ways in which your role/team can achieve this?
- Which way is most likely to have the largest impact on what your role/team is trying to achieve this?
- How will you validate or invalidate your hypotheses?

4. Data maturity level 2

At this level, you have a data infrastructure that gives you substantial data with sufficient accuracy.

Ideally, you have significant information about the people who click on your website. In essence, you have started to explain why things are happening using data.

You can start to define quantitative metrics and have clear goals and measures of the results of your work.

"If you can't measure it, you can't manage it"

Lagging metric - your ultimate goal but not one that you can directly work on (increase revenue, reduce costs, etc.)

Leading metric - a goal that you can work towards directly, which impacts the lagging metric (increase number of 'buy now' clicks)

What makes a good leading metric?

- **Movability** - your ability to move/influence the metric (whether you can impact the number of people clicking on 'buy now')
- **Stability** - a metric that is stable day to day
- **Non-gamability** - how easy is it to drive up the metric without actually working towards the end goal
- **Relevance** - how related is the leading metric to the lagging metric
- **Understandability** - is the metric clear to the entire team and leadership
- **Ability to measure** - a metric that is not confusing

Once you have defined a leading metric:

1. Brainstorm the projects that will impact the metric.
2. Do opportunity sizing of each project.
3. Choose the most promising projects

Measuring your progress towards your goal is important because you can:

1. double down on what appears to be working really well
2. make sure your metric is not changing due to something that you didn't do
3. make sure people get credit for their impact on the metric

3 stages you can work towards to establish a causal relationship between your work and the changes in your work:

- Crawl - analyse a trend before and after you make a change
- Walk ("Diff in Diff") - a quasi-experimental approach that compares the changes in outcomes over time between a population enrolled in a program and one that is not; for ex. you have two cohorts and offer a promotion to only one of them to see the impact of the promotion; The risk associated with this approach is that there may be naturally occurring differences between the two groups because after all they are different groups of people; A way to decrease this risk is to measure the metrics prior to running the experiment and confirm there is no difference between the two groups prior the experiment;
- Run - randomized control trial or A/B testing; You are able to randomize the control and test group; In this way, if there is a difference between your groups, you are much more confident that this is due to your treatment and not outside factors;

A pre-requisite for 'Run' state:

1. You need to be able to uniquely identify your target userbase
2. You must be able to randomize your target userbase and apply the same treatment
3. You need to be able to have a big enough sample size for each of your different groups to detect a difference between the subgroups

The run stage is considered 'the gold standard'.

When experimenting you should be careful about:

1. Deciding too soon

Novelty effect - when consumers respond to new things at a much higher rate than they would be responding after the phase is over.

2. Instability

3. Pull forward effect

For example, if a website does a sale of online courses on their site, they might see higher revenue at first, but since people will stock on courses, customers might not purchase additional materials for some time after the sale.

4. Occurrence of outliers

It is always important to inspect your data for data points that can skew it left or right.

5. Seasonality

Predictable changes in your metric like the day of the week, time of the day, etc.

6. Multi sample bias/ P-value hacking

If you take multiple different samples of your control group chances are that at least one of them will show statistically significant results.

4.1 Intro to Project 2

Business problem:

Sales for a new company have been sluggish.

To save the product line we must demonstrate that we can significantly increase our sales via marketing campaigns.

The solution needs to be:

- Incremental: result in sales that would not have happened otherwise
- Generate a meaningful volume of sales from a business perspective

What is our hypotheses?

We believe that emailing our customer base with information about the new toys will increase purchases.

We will consider a campaign successful if we drive +5% incremental sales within the first month.

How do you measure the outcomes?

- Crawl stage
- Walk stage
- Run stage

4.2 Crawl stage

Necessary skills:

- You have access to resources that can do some fundamental analytics
- Ability to understand and work with low volume data

Infrastructure (very limited):

- Test velocity is low
- You are unable to run multiple tests concurrently

Budget:

- No formal budget

How you measure impact?

You make your impact assessment by comparing how things were before you made a change (pre-period) to how things are after the change (post-period).

Risks:

- All other factors must remain the same - very hard to achieve

Quantifying impact:

Wait until your effect stabilizes (novelty effect needs to wear off)

Measure the difference between the original run-rate and the current run-rate

For example:

Daily average sales before change: 5

Daily average sales after change: 12

+140% increase in sales per day

4.3 Walk stage

Necessary skills:

You have access to resources that can do some fundamental analytics and understand testing fundamentals

Infrastructure:

- Capability to capture information about users and act on it by rolling out a marketing tactic

- You have historical data about customers. A period long enough to understand behaviour
- You can split customers into distinct and similar groups
- You can run multiple tests concurrently
- Test velocity is picking up but can be improved

Budget:

You have some financial support to scale successful discoveries to maximize impact

Quantifying impact:

- Pre-post analysis of groups (Diff-in-diff analysis)
- You measure the change in your target metric after you rollout your tactic
- You can make a statement like “the difference in average daily sales between the test and control groups was 500 before we revamped our website and 620 after”.
- You employ some crawl stage methods for areas that are yet to have split testing capabilities

Diff-in-diff: Measure the difference between a test and a control group and see if it is statistically significant

How to implement diff-in-diff:

- You want to leverage naturally occurring groups: cities
- Study historical data

4.4 Run stage

Necessary skills:

You have access to resources that can do some fundamental analytics and advanced modelling

Infrastructure:

- Capability to capture information about users and act on it by rolling out a marketing tactic
- You have robust in-house tools or strong integration into 3rd party tools to execute your testing and facilitate measurement
- Your test velocity is high and you can run multiple tests in parallel due to large data sets

Budget:

You have shown strong wins by leveraging data and as such you have solid financial support to discover and scale successful strategies.

Quantifying impact:

- You mainly use methods that confirm that action A leads to action B
- You use statistical measurement more rigorously to confirm your results
- You may still have cases where rigorous measurement is not possible

4.5 What is an A/B test

A test that proves or disproves a hypothesis around the causal relationship between an action you take and an outcome that can be observed and measured. Before investing in a test, do some background research to ground your hypothesis.

How to set up an A/B test

1. Identify your targets
2. Randomly split into two or more identical groups
3. Assign different treatments to each group and measure outcome

4.6 Statistical power

This metric quantifies your test's ability to detect an outcome when the outcome exists.

Typically set at 80%, this means that in 80 out of 100 tests where the effect exists, your test will indeed conclude that it exists.

In the crawl and walk stages, power is typically looked at as something 'nice-to-have' since what we're mostly looking for are directional signals about the impact of our tactics.

4.7 Sample size

All else being equal, the larger the sample size, the more powerful will be your test.

4.8 Test power vs lift

Example: We observe around 38% lift in purchase rate, but it is not statistically significant at the 95% level. We would want to explore a follow up test where we increase our sample size.

In this case we need to:

- Check for bias in Test vs Control groups
- Confirm if the effect has stabilized

Once you have demonstrated the value and potential impact of testing, you have effectively de-risked the decision-makers' call to invest in enabling and accelerating testing.

Armed with your current data and knowledge, find the next biggest initiative to invest in.

4.9 How to get to the next level

Two main ways to drive even more value to your business:

- Improve the quality of existing data: eliminating double logging of data, auditing and validating data, capturing events and transactions that may be missing

Without good data drawing insights from the data is much harder and the results may be misleading.

- Double down on testing

Consider building infrastructure; sampling and testing infrastructure would allow you to run experiments more quickly and efficiently. Work towards holding yourself to a higher statistical standard.

Holdouts: take a certain portion of the entire population and make sure they experience no treatments; then you go ahead with your project and apply treatments to the other part of the population.

Running experiments in parallel – make sure that the control and test groups are randomized across all the different experiments

4.10 How to ask for funding from decision makers

The framework is the same as the one discussed at Level 1. You want to:

- Demonstrate the impact
- Have an MVP of what you can do

Since now you are able to quantify your results on your metrics, you can report these obvious increases in the metrics and report statistically significant results.

Then you want to clearly illustrate how investing in:

- data talent
- engineering resources
- data infrastructure
- automation

would allow you to have:

- more accurate data
- move more efficiently
- drive even more impact

For example:

You discover that people who spend more than 10 minutes on your website respond particularly well to a specific discount. However, to do this you need to have real time logging on your website.

As a rule of thumb, investing in more robust datasets pays off. You are able to segment across different behaviour attributes and target individuals based on these attributes. You can even hire people to do machine learning, so you can do much better at automated targeting.

5. Data maturity level 3

By stage 3 you have a global understanding of cause-and-effect relationships. You know who is likely to buy what product given what kind of treatment and scenarios there are. For example, you will know that if you give people \$10 discounts you will increase profits more than if you give them \$15 or \$8 discounts.

The data you may be tracking:

- age
- gender
- where people click on the website
- location

All of this depends on the legal landscape in your country. Make sure you stay true to your values when collecting data.

Rule of thumb: If you don't want other people tracking something about you, chances are your customers would not want you to track them either.

From a growth perspective, at this stage you have likely run out of low risk - high reward ideas to run A/B tests on.

Infrastructure and data quality

- Robust infrastructure for storing and querying data
- Trustworthy data

Ideally, you would have checkpoints in place to make sure that new data satisfies specific quality standards. You also have a team of engineers and data people maintaining infrastructure and data quality.

Automation

- Many known impactful solutions are automated.

- Data engineers monitor automated functions
- Experimental framework:

Parallel experimentation:

- Randomization
- Parallel experimentation
- Population targeting
- Sample sizes

Not all areas of the company are automated at stage 3 data maturity. Using your data team's experience and accomplishments, you can help scale your solutions across other areas of the company. For example, scaling your automation efforts for providing discounts to maximize revenue for other products and services. Or you can build template dashboards for people across the company to visualize and track metrics more easily.

At this level, you also likely have the talent and robust data to explore advanced algorithms like machine learning algorithms to identify more complex second order relationships. For example, you can use such algorithms for:

- Sales forecasting
- Anomaly detection
- Optimize data storage and energy usage
- Targeting models
- Recommender models

Advanced algorithms are beneficial to a company only when it reaches level 3 maturity.

6. What to do to improve after level 3

There is not a clearly defined higher data maturity level that is widely adopted in industry.

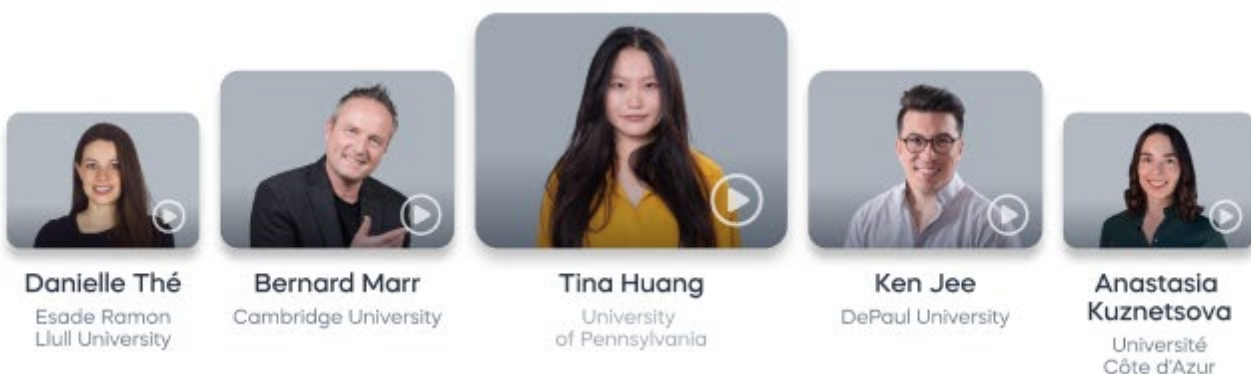
Some areas of focus can be:

1. Automating all known solutions and features
2. Build more complex models to automate routine activities
3. Identify new stage 0 opportunities in your business and restart the process with all the experience you have
4. Seek to apply innovations (innovations falling in the following categories:
 - personalize - recommendations based on what a customer or a potential customer is interested in
 - act faster - gives you more context into your space
 - discover hidden interactions -
 - better infrastructure and scaling

Learn DATA SCIENCE anytime, anywhere, at your own pace.

If you found this resource useful, check out our **e-learning program**. We have everything you need to succeed in data science.

Learn the most sought-after data science skills from **the best experts in the field!**
Earn a **verifiable certificate** of achievement trusted by employers worldwide and future proof your career.



Comprehensive training, exams, certificates.

- ✓ 162 hours of video
- ✓ 599+ Exercises
- ✓ Downloadables
- ✓ Exams & Certification
- ✓ Personalized support
- ✓ Resume Builder & Feedback
- ✓ Portfolio advice
- ✓ New content
- ✓ Career tracks

Join a global community of 1.8 M successful students with an annual subscription
at 60% OFF with coupon code **365RESOURCES**.

~~\$432~~ **\$172.80**/year



Start at 60% Off

VAT may be applied

365[✓]DataScience



Davis Balaba

Email: team@365datascience.com

365[✓]DataScience