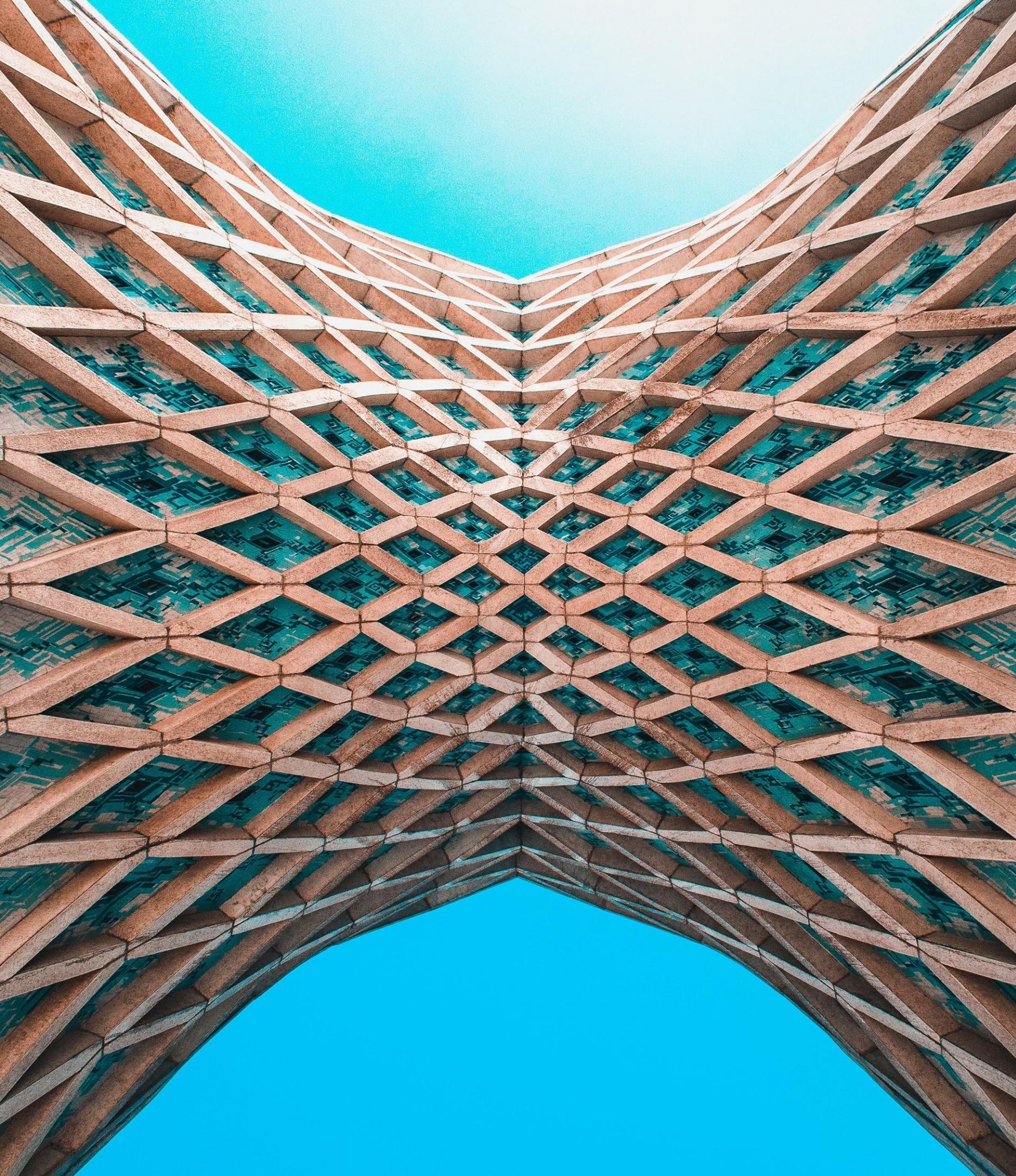


Data Science Project Design

Objectives

Principles of good project design:

- Focus on the user
- Start simple
- Iterate
- Practice design thinking



Human-Centered Design

One Approach to Data Science

1. Start with some data
2. Find out what problems your data can address
3. Do analysis
4. Present the results

One Approach to Data Science

1. Start with some data
2. Find out what problem your data can add
3. Do analysis to get results
4. THERE'S A BETTER WAY!

One Approach to Data Science

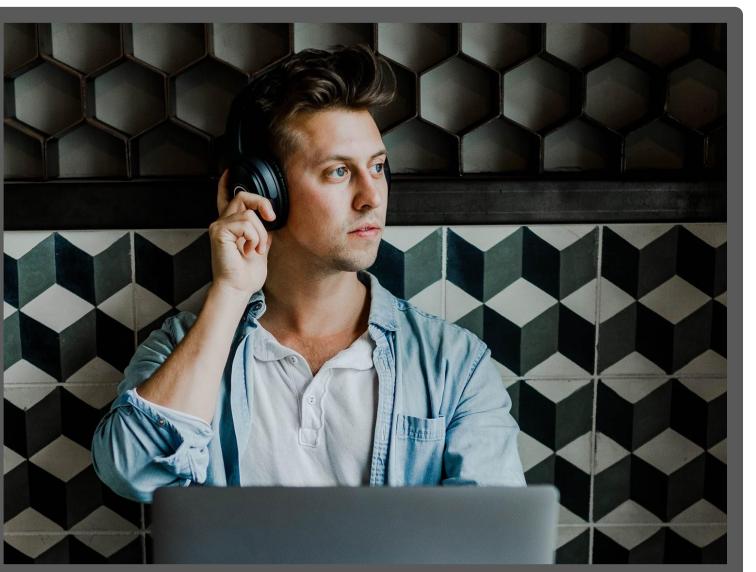
1. Start with some data

**2. Find out what problems your
data can address**

3. Do analysis

4. Present the results

Start with the End User



Start with the End User

Who will the project help?

What problems does the end user face?

How does this project address those problems?

Example of Human-Centered Design

Who will the project help?

What problems does the end user face?

How does this project address those problems?

Example of Human-Centered Design

Who will the project help?

Adults learning a new language

What problems does the end user face?

How does this project address those problems?

Example of Human-Centered Design

Who will the project help?

Adults learning a new language

What problems does the end user face?

Limited practice time; Forget to study

How does this project address those problems?

Example of Human-Centered Design

Who will the project help?

Adults learning a new language

What problems does the end user face?

Limited practice time; Forget to study

How does this project address those problems?

App with daily push notifications that pop up when user is less busy

A Better Data Science Approach

1. Start with the **end user**

- Human-centered design
- Empathy
- Focus on end user and goal instead of modeling

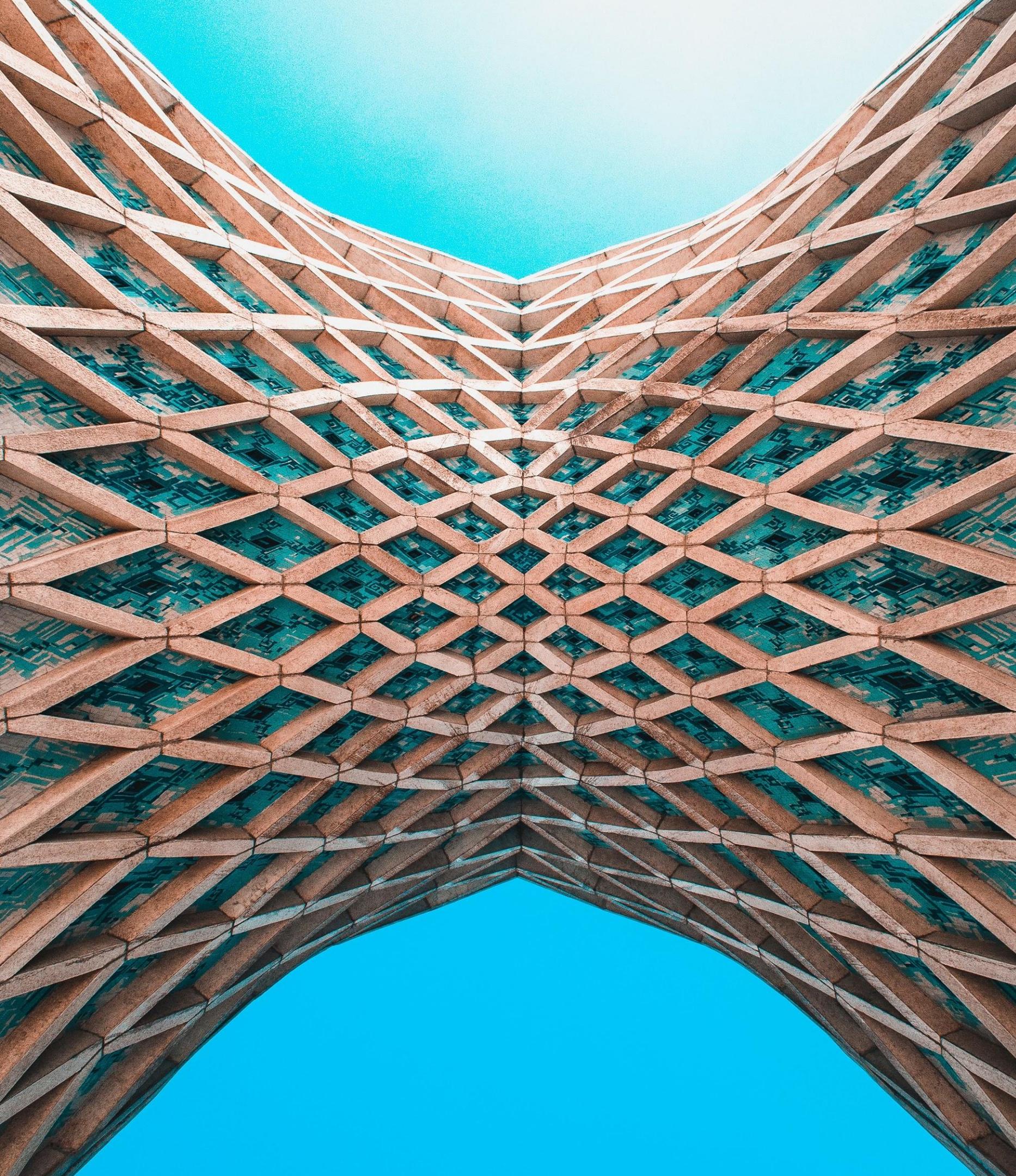


A Better Data Science Approach

1. Start with the **end user**
2. **Brainstorm** ideas based on problems the end user faces and possible ways to help

A Better Data Science Approach

1. Start with the **end user**
2. **Brainstorm** ideas based on problems the end user faces and possible ways to help
3. Prototype ideas to create **Minimum Viable Product**
4. **Iterate**



Minimum Viable Product (MVP)

A Better Data Science Approach

1. Start with the **end user**
2. **Brainstorm** ideas based on problems the end user faces and possible ways to help
3. Prototype ideas to create **Minimum Viable Product**
4. **Iterate**

A Better Data Science Approach

1. Start with the end user
2. Brainstorm ideas based on problems the end user faces and possible ways to help
- 3. Prototype ideas to create Minimum Viable Product (MVP)**
4. Iterate

What is an MVP?

- Minimum Viable Product
- First prototype solution
- Created quickly
- Addresses project goals

What makes a good MVP?

Minimum

Could something simpler or quicker accomplish a similar outcome?

What makes a good MVP?

Minimum - quick and simple

What makes a good MVP?

Minimum - quick and simple

Viable

Does the MVP work?

What makes a good MVP?

Minimum - quick and simple

Viable - works

What makes a good MVP?

Minimum - quick and simple

Viable - works

Product

Does the MVP address a need of the audience and add value?

What makes a good MVP?

Minimum - quick and simple

Viable - works

Product - adds value

Example MVP

Problem Statement

Adults learning new language need study reminders to fit into their schedules

Example MVP

Problem Statement

Adults learning new language need study reminders to fit into their schedules

Proposed Project

App with daily push notifications that pop up when user is less busy

Example MVP

Problem Statement

Adults learning new language need study reminders to fit into their schedules

Proposed MVP

App with daily push notifications that pop up when user is less busy

Example MVP

Problem Statement

Adults learning new language need study reminders to fit into their schedules

Proposed MVP

Daily emails

~~App with daily push notifications~~
that pop up when user is less busy

Example MVP

Problem Statement

Adults learning new language need study reminders to fit into their schedules

Proposed MVP

Daily emails

~~App with daily push notifications that pop up when user is less busy~~
that arrive in the morning or at night based on past study times

Example MVP

Proposed MVP

Daily emails that arrive in the morning or at night based on past study times

- ✓ **Minimum** - quick and simple
- ✓ **Viable** - works
- ✓ **Product** - adds value

Why start with an MVP?

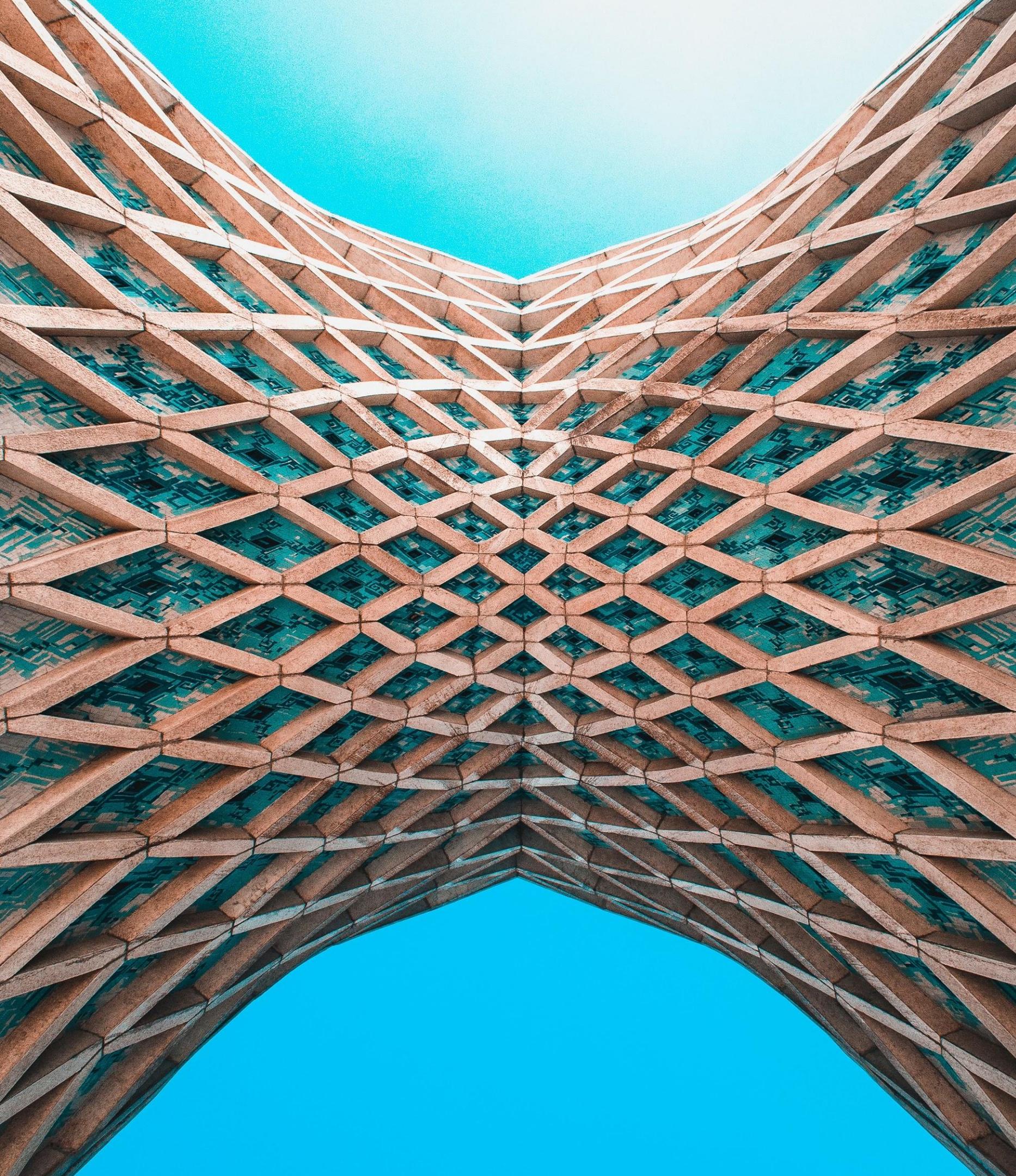
- Working solution available sooner
- Ask for feedback
- Primary goal of MVP is to **learn**

MVP Recap

Minimum - quick and simple

Viable - works

Product - adds value



Iterative Design Process

A Better Data Science Approach

1. Start with the **end user**
2. **Brainstorm** ideas based on problems the end user faces and possible ways to help
3. Prototype ideas to create **Minimum Viable Product**
4. **Iterate**

A Better Data Science Approach

1. Start with the end user
2. Brainstorm ideas based on problems the end user faces and possible ways to help
3. Prototype ideas to create Minimum Viable Product
- 4. Iterate**

Iterative Design Process #1

Client Request

“I need a car to get from point A to point B.”

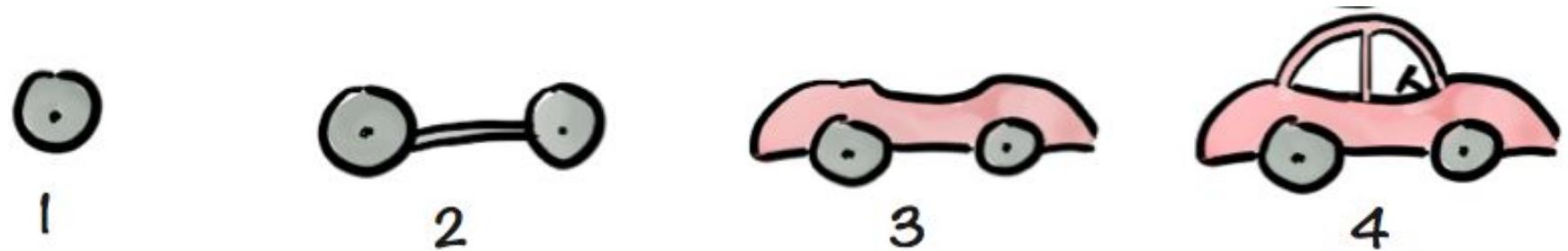
Drawing and Concept by Henrik Kniberg

Iterative Design Process #1

Client Request

“I need a car to get from point A to point B.”

Your Project Plan



Drawing and Concept by Henrik Kniberg

Iterative Design Process #1

First Meeting



I

It's not
a car.

Iterative Design Process #1

First Meeting



1

It's not
a car.

Second and Third Meetings



2



3

Not. A.
Car.

Iterative Design Process #1

Final Meeting



Iterative Design Process #2

Client Request

“I need a car to get from point A to point B.”

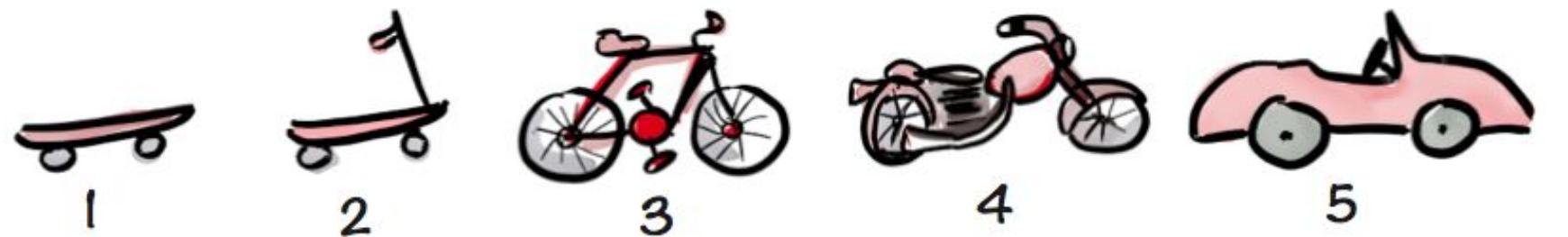
Drawing and Concept by Henrik Kniberg

Iterative Design Process #2

Client Request

“I need a car to get from point A to point B.”

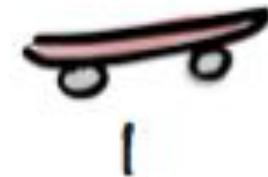
Your New Project Plan



Drawing and Concept by Henrik Kniberg

Iterative Design Process #2

First Meeting



It's not a car.
I'm falling over.

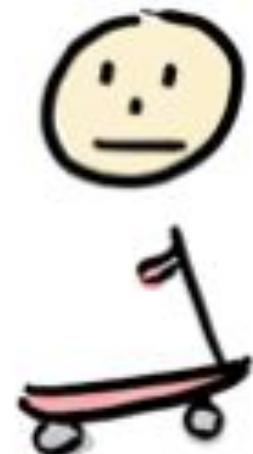
Iterative Design Process #2

First Meeting



It's not a car.
I'm falling over.

Second Meeting

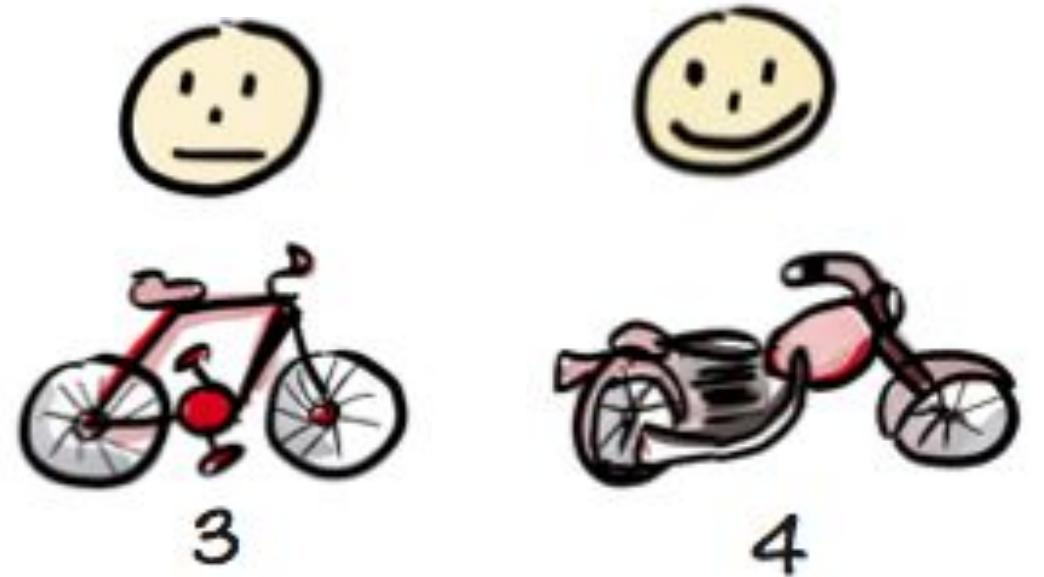


2

Good. Not
falling over.
But I can't go
very far.

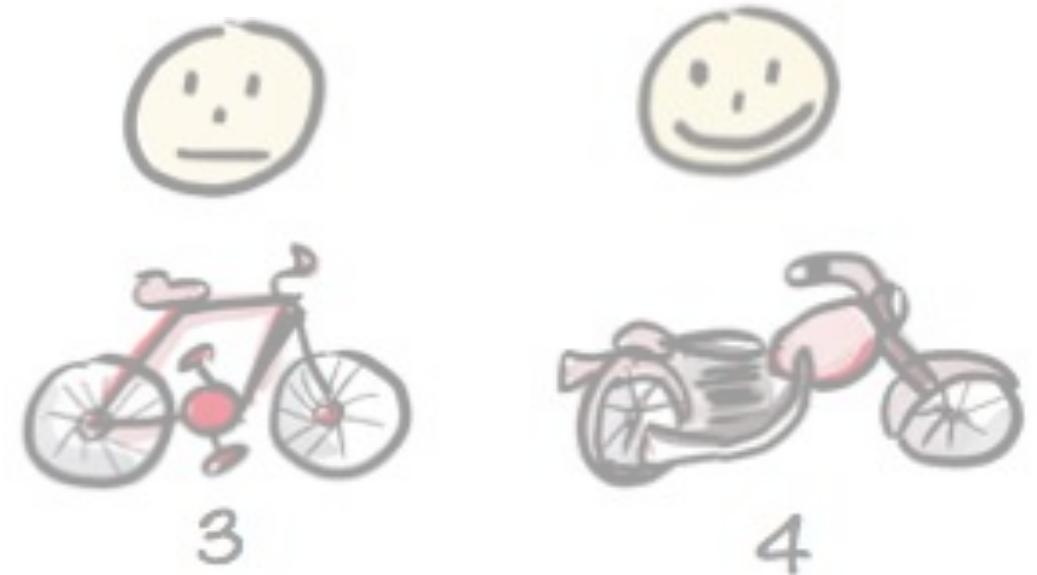
Iterative Design Process #2

Third and Fourth Meetings



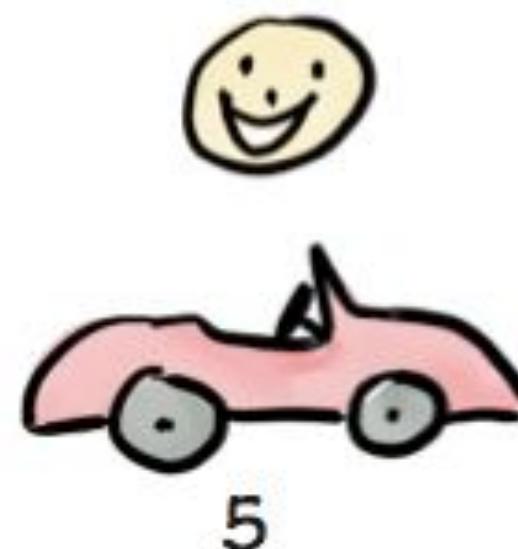
Iterative Design Process #2

Third and Fourth Meetings



Getting much
better and I
love the wind in
my face!

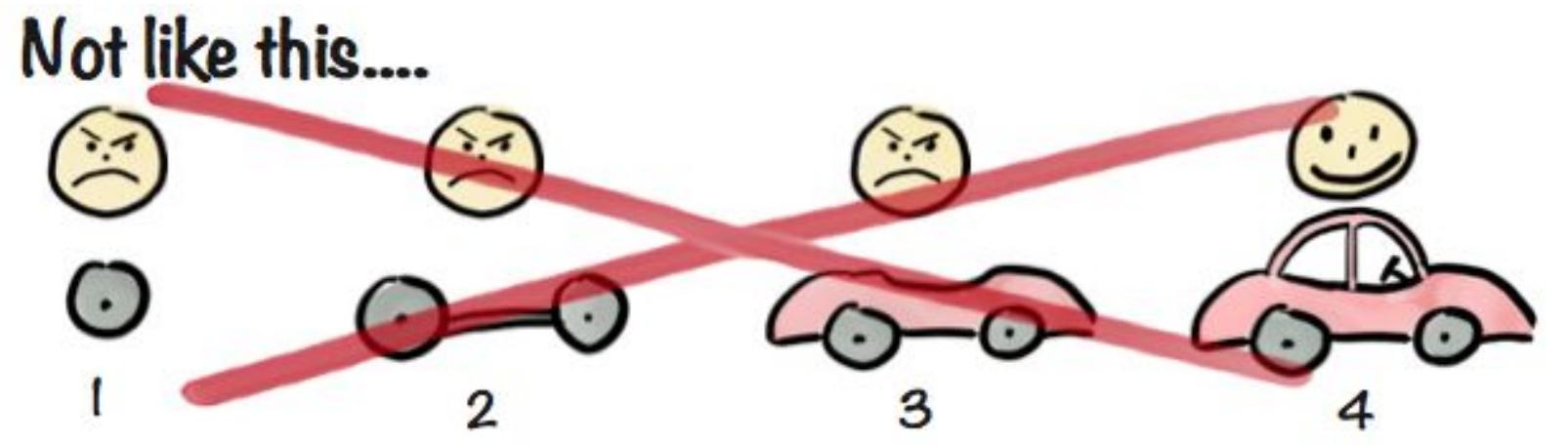
Final Meeting



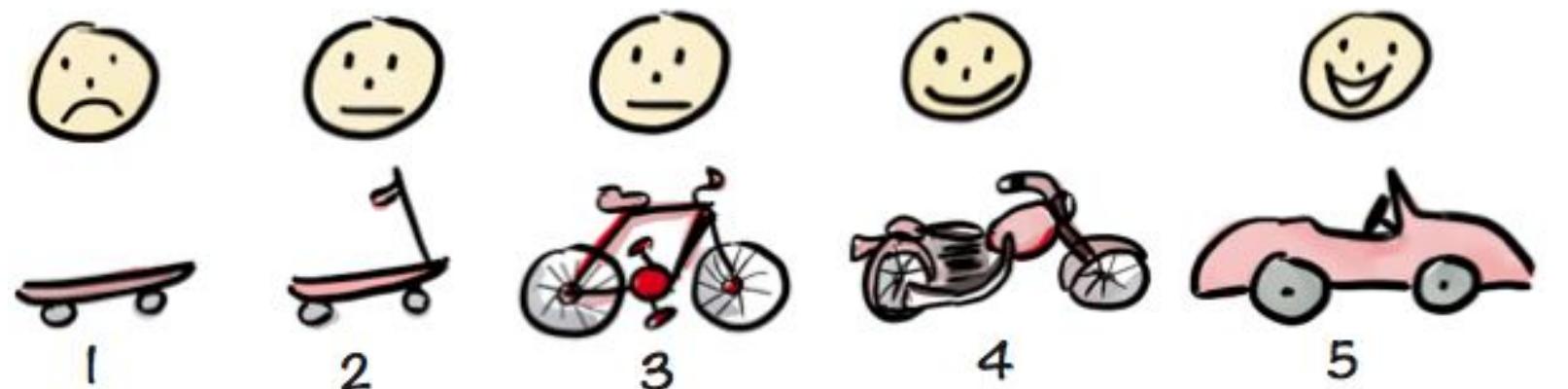
Iterative Design Process

Client Request

"I need a car to get from point A to point B."



Like this!



Drawing and Concept by Henrik Kniberg

Principles of Good Design

- **Define MVP** before data collection, modeling, etc.
 - Start with a simple, working solution
 - Ask for feedback

Principles of Good Design

- **Define MVP** before data collection, modeling, etc.
- Build each step like you're **planning to iterate** it
 - Clear thought process
 - Comments/documentation
 - Functional better than perfect

Plan to Iterate on Each Step



Data Acquisition	One week of data from .csv
Data Cleaning	Lowercase words, English Stopwords
Model	Basic Naive Bayes
...	...
Output	Text document describing goals, results

Plan to Iterate on Each Step



Data Acquisition	One week of data from .csv	One month of data from API
Data Cleaning	Lowercase words, English Stopwords	Custom stopwords, links removed
Model	Basic Naive Bayes	SVD, Logistic Regression
...
Output	Text document describing goals, results	Presentation slides with main points

Plan to Iterate on Each Step



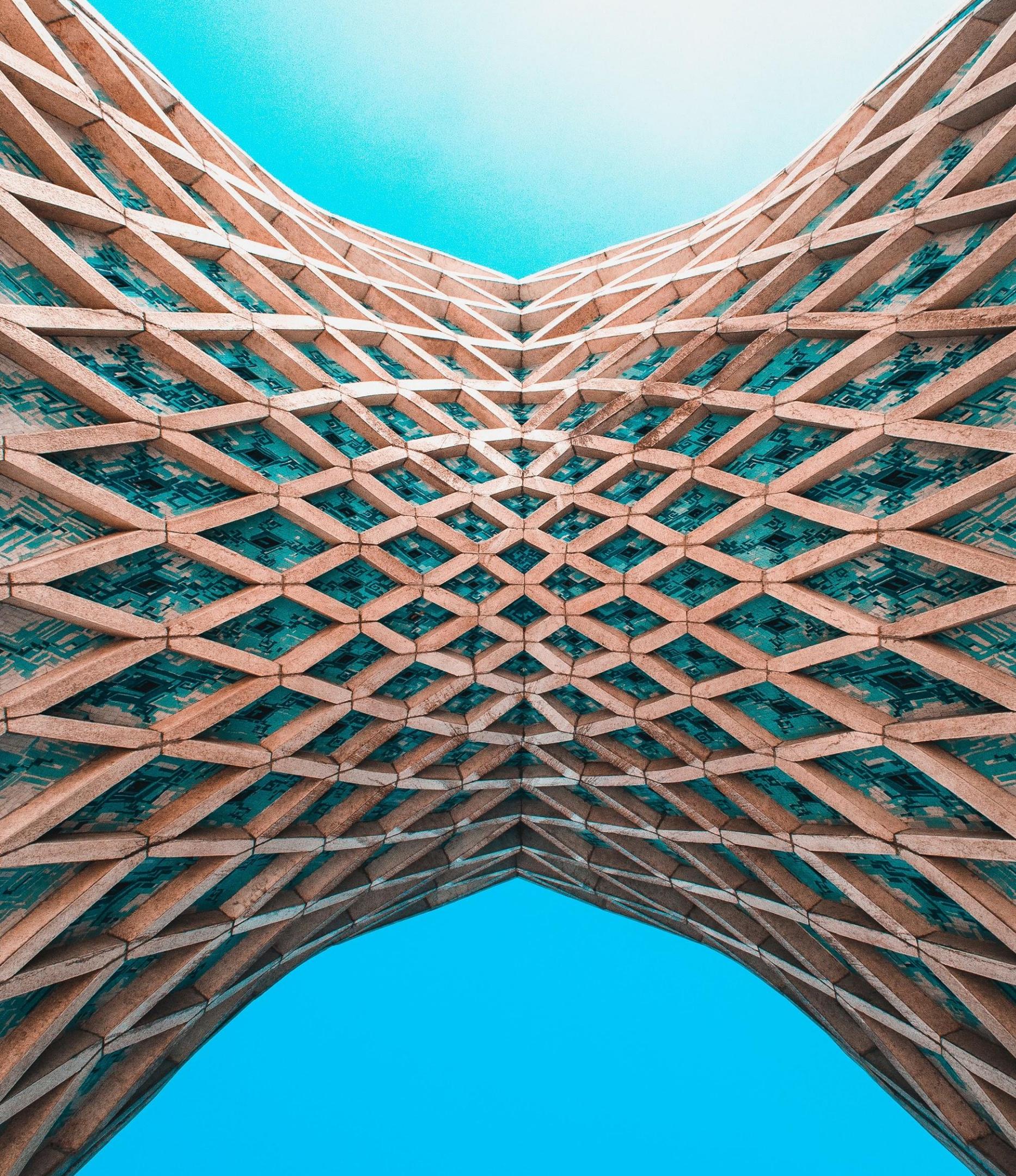
Data Acquisition	One week of data from .csv	One month of data from API	...	Text merged with API metadata, Scraped images
Data Cleaning	Lowercase words, English Stopwords	Custom stopwords, links removed	...	Parts of speech, Images filters
Model	Basic Naive Bayes	SVD, Logistic Regression	...	Recurrent NN / Convolutional NN
...
Output	Text document describing goals, results	Presentation slides with main points	...	Flask app with filters, sliders, images of recos

Principles of Good Design

- **Define MVP** before data collection, modeling, etc.
- Build each step like you're **planning to iterate** it
- Use later steps to **inform decisions** about earlier steps
 - Model may guide data scraping
 - Visualization dashboard may need specific data cleaning

Principles of Good Design

- **Define MVP** before data collection, modeling, etc.
- Build each step like you're **planning to iterate** it
- Use later steps to **inform decisions** about earlier steps



Applications of the Design Process

A Better Data Science Approach

1. Start with the **end user**
2. **Brainstorm** ideas based on problems the end user faces and possible ways to help
3. Prototype ideas to create **Minimum Viable Product**
4. **Iterate**

Example #1 - Student Project

Problem Statement

A grocery store delivery company needs help planning product demand for inventory purposes



Example #1 - Student Project

Problem Statement

A grocery store delivery company needs help planning product demand for inventory purposes

1. Who is the **end user**?

- Stakeholder clients at grocery delivery company
- Class presentation

Example #1 - Student Project

Problem Statement

A grocery store delivery company needs help planning product demand for inventory purposes

2. **Brainstorm** how to help client

- Predict if customer will repurchase specific items in their next delivery
- Determine expected amount and size of new customer orders

Example #1 - Student Project

Problem Statement

A grocery store delivery company needs help planning product demand for inventory purposes

2. Brainstorm how to help client

- **Predict if customer will repurchase specific items in their next delivery**
- Determine expected number and size of new customer orders

Example #1 - Student MVP

3. Define MVP and build



- Data from subset of past customers
- Logistic regression model
- One feature: number of previous customer-product purchases
- One-page summary with metrics and a few plots

Example #1 - Student Iterations



4. **Iterate** based on feedback

- More customers in database
- More sophisticated classification model (*random forest, boosting*)
- Do more feature engineering
- Mock up client dashboard or app

Example #2 - Company Project

Problem Statement

Customers are visiting e-commerce website but not purchasing



Example #2 - Company Project

Problem Statement

Customers are visiting e-commerce website but not purchasing

1. Who is the **end user**?

- Jewelry customers
- Tech-savvy
- Shoppers in their 20s and 30s

Example #2 - Company Project

Problem Statement

Customers are visiting e-commerce website but not purchasing

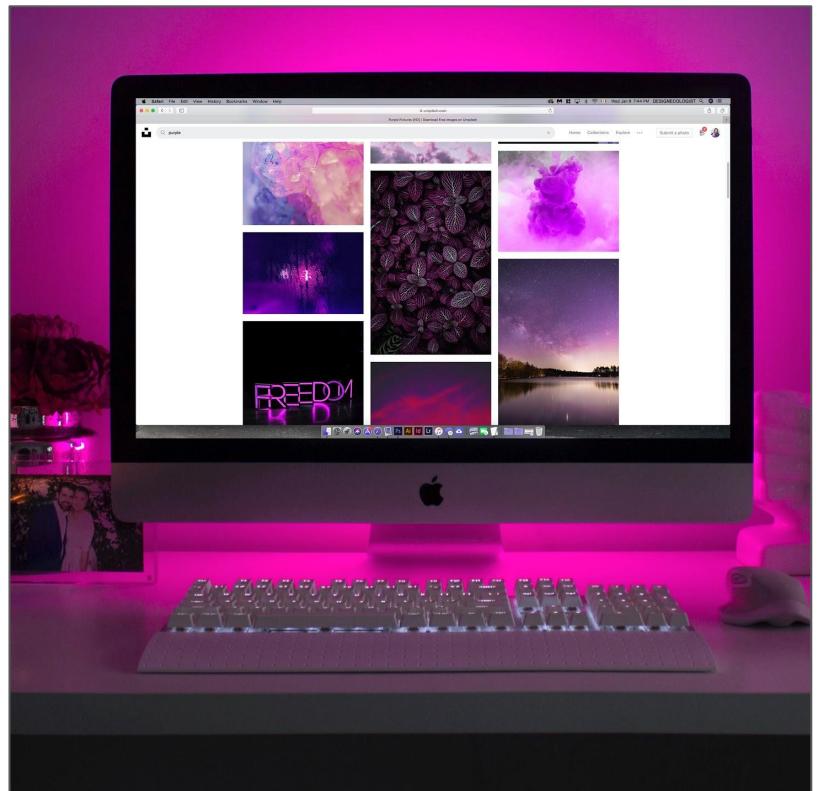
2. **Brainstorm** how to help users

- Conduct customer interviews
- Analyze reviews
- Good quality products, hard time finding items of interest

Example #2 - Company Project

Possible improvements:

Site Filters



Product
Recommendations



Agile Project Management

MVP and Iterate

=

Agile Project Management

- Lifecycles with incremental improvements
- Scope defined before each iteration
- Scrum: 2-week sprints

Each Agile Sprint

- Team members responsible for specific tasks
- Project manager tracks progress
- Daily stand ups to identify blockers
- Retrospective at end of each sprint
- Iterative improvements and adaptive to feedback

A Better Data Science Approach

1. Start with the **end user**
2. **Brainstorm** ideas based on problems the end user faces and possible ways to help
3. Prototype ideas to create **Minimum Viable Product**
4. **Iterate**