

Data\_Driven\_Design



Contents

# Starting Points

## Accessing Tools

### Download KNIME

<https://www.knime.com/downloads>

### Download Power BI Desktop

<https://powerbi.microsoft.com/en-us/downloads/>

## Accessing Course Materials

### Access this Mindmap

<https://www.mindmeister.com/1962731121/data-driven-design>

#### https://bit.ly/SMU\_DDD

### Copy this Mindmap

<https://academy.smu.edu.sg/node/646>

## Recapping Course Objectives

<https://academy.smu.edu.sg/data-analytics-using-power-bi-2636>

# Framing

## ...seeding some important perspectives...

### Design

#### ...design is a process and different process frameworks have developed to address different design challenges...

##### Design Processes

<https://www.designthinkersgroup.us/wp-content/uploads/2020/10/Picture1-e1604085814680.png>

###### ...more...

Design Thinkers Group

<https://www.designthinkersgroup.us/design-thinking-vs-lean-startup-vs-agile-vs-six-sigma/>

#### ...there are two very dominant viewpoints in this spectrum...

##### Design Thinking for Service Design

[https://images.squarespace-cdn.com/content/v1/57c6b79629687fde090a0fdd/1555537929547-MZ796SNQ3Z1B983VJFIR/ke17ZwdGBToddI8pDm48kL8n1BY0dxneYBSjNw5Xpt9Zw-zPPgdn4jUwVcJE1ZvWQUxwkmyExglNqGp0IvTJZamWLI2zvYWH8K3-s\_4yszcp2ryTI0HqTOaaUohrI8PIpbkVoMq2LdeqYkZJMf8nBnGIWAnafX0ITf5Jdd-nK3s/Stanford\_dschool\_ProcessHexDiagram\_Tool\_Behaviors+2019](https://images.squarespace-cdn.com/content/v1/57c6b79629687fde090a0fdd/1555537929547-MZ796SNQ3Z1B983VJFIR/ke17ZwdGBToddI8pDm48kL8n1BY0dxneYBSjNw5Xpt9Zw-zPPgdn4jUwVcJE1ZvWQUxwkmyExglNqGp0IvTJZamWLI2zvYWH8K3-s_4yszcp2ryTI0HqTOaaUohrI8PIpbkVoMq2LdeqYkZJMf8nBnGIWAnafX0ITf5Jdd-nK3s/Stanford_dschool_ProcessHexDiagram_Tool_Behaviors%2B2019)

###### ...more...

The Design Thinking Process

<https://www.ideou.com/pages/design-thinking#process>

##### Six Sigma for Process Design

<https://goproductivity.ca/wp-content/uploads/DMAIC.png>

#### ...but design is increasingly "data-driven" in a digital economy and this has eroded tribal boundaries...

##### Kayak A-B Tests

<https://apptimize.com/blog/2014/03/kayaks-most-interesting-ab-test/>

##### German Rail

<https://www.contagious.com/news-and-views/german-rail-campaign-facebook-sheryl-sandberg-called-the-future-of-advertising>

##### Triggr Health

<https://www.vice.com/en_us/article/7x4ewz/i-tried-an-app-meant-to-disrupt-alcoholism>

#### ...we can see some interesting fusions of "hard" and "soft" design approaches in leading organisations...

##### "Uber Engineering is committed to developing technologies that create seamless, impactful experiences for our customers..."

###### Uber Michelangelo Design Outcomes

<https://eng.uber.com/scaling-michelangelo/>

...more...

Uber Michelangelo

<https://eng.uber.com/michelangelo-machine-learning-platform/>

#### ...in summary we can look to complexity theory to guide us through the selection of specific design techniques for specific contexts...

##### The Cynefin Model

<https://www.rochester.edu/aincenter/wp-content/uploads/2019/02/Cynefin-Framework.png>

###### ...more...

David Snowden and Cognitive Edge

<https://www.youtube.com/watch?v=N7oz366X0-8>

### Two Views

#### ...throughout the course we are going to explore the similarities and differences between two design approaches...

##### Design Thinking

<https://media.nngroup.com/media/editor/2016/07/29/designthinking_illustration_final-01-01.png>

###### ...steps...

Empathise

...detail...

Conduct research in order to develop knowledge about what your users do, say, think, and feel.

Define

...detail...

Combine all your research and observe where your users’ problems exist. In pinpointing your users’ needs, begin to highlight opportunities for innovation.

Ideate

...detail...

Brainstorm a range of crazy, creative ideas that address the unmet user needs identified in the define phase. Give yourself and your team total freedom; no idea is too farfetched and quantity supersedes quality.

Prototype

...detail...

Build real, tactile representations for a subset of your ideas. The goal of this phase is to understand what components of your ideas work, and which do not. In this phase you begin to weigh the impact vs. feasibility of your ideas through feedback on your prototypes.

Test:

...detail...

Return to your users for feedback. Ask yourself ‘Does this solution meet users’ needs?’ and ‘Has it improved how they feel, think, or do their tasks?’

Implement

...detail...

Put the vision into effect. Ensure that your solution is materialized and touches the lives of your end users.

###### ...guides...

d.school Bootleg

<https://dschool.stanford.edu/resources/design-thinking-bootleg>

IDEO Design Thinking Process

<https://www.ideou.com/pages/design-thinking#process>

Nielsen Norman Group

<https://www.nngroup.com/articles/design-thinking/>

IDEO on Design Thinking

<https://designthinking.ideo.com/>

Design Kit Case Studies

<https://www.designkit.org/case-studies/3>

##### CRISP

<https://www.tibco.com/sites/tibco/files/media_entity/2019-03/data_science-diagram.svg>

###### ...steps...

Business Understanding

...detail...

This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

Data Understanding

...detail...

The data understanding phase starts with initial data collection and proceeds with activities that enable you to become familiar with the data, identify data quality problems, discover first insights into the data, and/or detect interesting subsets to form hypotheses regarding hidden information.

Data Preparation

...detail...

The data preparation phase covers all activities needed to construct the final dataset [data that will be fed into the modeling tool(s)] from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools.

Modelling

...detail...

In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, going back to the data preparation phase is often necessary.

Evaluation

...detail...

At this stage in the project, you have built a model (or models) that appears to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to thoroughly evaluate it and review the steps executed to create it, to be certain the model properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

Deployment

...detail...

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. It often involves applying “live” models within an organization’s decision making processe —for example, real-time personalization of Web pages or repeated scoring of marketing databases. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases, it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will carry out the deployment effort, it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models.

###### ...guides...

Uber Michelangelo Design Process

<https://1fykyq3mdn5r21tpna3wkdyi-wpengine.netdna-ssl.com/wp-content/uploads/2018/11/image6.png>

Tibco CRISP

<https://www.tibco.com/reference-center/what-is-data-science>

CRISP-DM Standard

<https://www.the-modeling-agency.com/crisp-dm.pdf>

# Phases of Design

## ...getting to grips with tools and techniques...

### Insight

#### ...the objective of the "Insight" phase is to develop "Systemic Understanding"...

##### Design Thinking

<https://media.nngroup.com/media/editor/2016/07/29/designthinking_illustration_final-01-01.png>

###### ...the design thinking method focusses attention on empathy...

Experiential Learning and Goat Man

<https://www.youtube.com/watch?v=VR5hFxeBaFY>

...empathy is good in the experience economy...

The Experience Economy

<https://www2.deloitte.com/au/en/pages/technology/articles/experience-transformation-through-better-customer-experience-management.html>

...journey mapping is a principal form of ethnographic research...

A journey map is an infographic visualization of the process that a persona segment goes through in order to accomplish a goal. Journey maps are useful in communicating the general narratives and themes uncovered by longitudinal research done to understand how a customer works toward a goal over time...

...example...

DBS Digibank

[https://medium.com/@kenleesanekata/enhancing-the-dbs-singapore-digibank-credit-card-payment-experience-a78467bd010d](https://medium.com/%40kenleesanekata/enhancing-the-dbs-singapore-digibank-credit-card-payment-experience-a78467bd010d)

Journey Map Application

<https://www.nngroup.com/articles/analyze-customer-journey-map/>

...more...

Journey Map Overview

<https://www.nngroup.com/videos/journey-mapping-101/>

Files: [fileIdscreen-shot-2018-11-06-at-122254-pm.png](https://www.mindmeister.com/generic_files/get_attachment/12660099)

Journey Map Guide

<https://www.nngroup.com/articles/journey-mapping-101/>

Journey Map Template

<https://media.nngroup.com/media/articles/attachments/JMTemplate.pdf>

###### ...insights from the research are used to form a design challenge statement...

A design challenge articulates the problem you are trying to solve, and helps you define a scope that is neither too narrow nor too broad...

...example...

"How Might We...?" Questions

<https://www.nngroup.com/articles/how-might-we-questions/>

...more...

Design Kit Design Challenges

<https://www.designkit.org/methods/frame-your-design-challenge>

How Might We Guide

<https://medium.com/archipelagos/how-might-we-f618c041c107>

Design Kit HMWs

<https://www.designkit.org/methods/3>

##### Data Science

###### ...the data scientist will typically pursue "Features", "Patterns", and "Models"...

Feature

...a feature is an event or condition that has an impact on the outcome we are studying...

CRISP-DM

<https://www.tibco.com/sites/tibco/files/media_entity/2019-03/data_science-diagram.svg>

...more...

What is Data Science?

<https://www.tibco.com/reference-center/what-is-data-science>

...sometimes we have to put in some pre-work to find features...

System and Concept Maps

<https://okrexamples.co/>

...more...

Ishikawa Diagram

<https://en.wikipedia.org/wiki/Ishikawa_diagram>

Reality Charts by Think Reliability

<https://www.thinkreliability.com/root-cause-analysis-tools/cause-mapping-examples/>

Exploratory Data Analysis Workflows

<https://devopedia.org/images/article/75/1899.1595429899.png>

...more...

EDA Overview

<https://devopedia.org/exploratory-data-analysis>

Standard EDA Framework

Files: [fileIdedamethods.png](https://www.mindmeister.com/generic_files/get_attachment/11546059)

Pattern

...a pattern reveals the nature of the relationship between the feature and the outcome...

Anscombe's Quartet

<https://upload.wikimedia.org/wikipedia/commons/thumb/e/ec/Anscombe%27s_quartet_3.svg/1280px-Anscombe%27s_quartet_3.svg.png>

...more...

Anscombe's Quartet

<https://en.wikipedia.org/wiki/Anscombe%27s_quartet>

A Brief History of the Scatter Plot - Data Visualization’s Greatest Invention

<https://qz.com/1235712/the-origins-of-the-scatter-plot-data-visualizations-greatest-invention/#:~:text=Originally%20called%20scatter%20diagrams%2C%20scatter,lines%20representing%20the%20two%20axes.>

...data visualisation techniques are particularly accomplished when it comes to revealling pattern...

John Snow and Cholera in 1854

<https://upload.wikimedia.org/wikipedia/commons/thumb/2/27/Snow-cholera-map-1.jpg/1024px-Snow-cholera-map-1.jpg>

...more...

1854 Broad Street Cholera Outbreak

<https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak>

David McCandless

<https://www.youtube.com/watch?v=5Zg-C8AAIGg>

...more...

Information is Beautiful

<https://informationisbeautiful.net/>

10 Rules for Better Dashboard Design

<https://uxplanet.org/10-rules-for-better-dashboard-design-ef68189d734c>

Chart Selection Frameworks

Files: [fileId1\_9VanG02d4If1TOIbpcTWbA.jpeg](https://www.mindmeister.com/generic_files/get_attachment/11545954), [fileIdchart-selection-diagram.png](https://www.mindmeister.com/generic_files/get_attachment/13470547), [fileId1\_dp7hLF\_b\_jrrM5h5aAutjQ.jpeg](https://www.mindmeister.com/generic_files/get_attachment/13470548)

...we can interpret visualisations in terms of feature and outcome...

Covid Vaccines

<https://cdn.statcdn.com/Infographic/images/normal/23510.jpeg>

...more...

How Effective Are The Covid-19 Vaccines?

<https://www.statista.com/chart/23510/estimated-effectiveness-of-covid-19-vaccine-candidates/>

Florida Gun Deaths

<https://cdn.mos.cms.futurecdn.net/h5MSdPM97fm55kTk4kk4P7-1024-80.jpg.webp>

...more...

Misleading Gun-Death Chart Draws Fire

<https://www.livescience.com/45083-misleading-gun-death-chart.html>

...but we have to take care to consider whether the patterns show correlation or causation...

Spurious Correlations

<https://www.tylervigen.com/spurious-correlations>

Model

...analysis of feature and pattern enables us to craft models that capture the essence of the system we are studying...

y = (f)x + e

<https://images.squarespace-cdn.com/content/v1/5d36db741438f10001b5027b/1565706752115-B5YFRFMP7KGMG2CNLI88/yfx.png>

...more...

How to Use Six Sigma’s Breakthrough Equation: Y = f(X) + ε

<https://www.dummies.com/careers/project-management/six-sigma/how-to-use-six-sigmas-breakthrough-equation-y-fx/>

...the first step for many modelling exercises will be the creation of a regression model...

Regression Fundamentals

<https://docs.google.com/drawings/d/1QovfMtB3tXIXrF5wzvVbEHf-QC2861Cbhxek-SFakso/edit?pli=1>

...more...

Understanding Regression Error Metrics

<http://www.pybloggers.com/2018/09/understanding-regression-error-metrics/>

Regressions can Classify or Predict

[https://miro.medium.com/max/1400/1\*G3imr4PVeU1SPSsZLW9ghA.png](https://miro.medium.com/max/1400/1%2AG3imr4PVeU1SPSsZLW9ghA.png)

...more...

A Comparison between Linear and Logistic Regression

[https://medium.com/@maithilijoshi6/a-comparison-between-linear-and-logistic-regression-8aea40867e2d](https://medium.com/%40maithilijoshi6/a-comparison-between-linear-and-logistic-regression-8aea40867e2d)

Regression for Values (Prediction)

<https://usedcarpricepredictor.herokuapp.com/>

Regression for Events (Classification)

<https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=3446&context=sis_research>

...more...

Predict an Event or Classify a Case

<https://www2.deloitte.com/us/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html/>

###### ...we can explore and develop these concepts in the following scenarios...

In Excel

...framing...

Q: Can I Predict and Control Commuting Behaviour in a City?

Our Data

Files: [fileIdTransport.xlsx](https://www.mindmeister.com/generic_files/get_attachment/11546020)

...doing...

Feature

...identifying data that can represent "Commuter Behaviour"...

An observable metric that can reliably represent the phenomenon

...identifying data that could influence "Commuter Behaviour"...

Events or conditions that we may suspect are influential in this system (these are hypotheses, we may be wrong)

...more...

Design an Effective Hypothesis

<https://help.optimizely.com/Ideate_and_Hypothesize/Design_an_effective_hypothesis>

Pattern

...confirming and exploring relationships between outcome and feature...

Create Four Scatter Charts

<https://support.microsoft.com/en-us/topic/present-your-data-in-a-scatter-chart-or-a-line-chart-4570a80f-599a-4d6b-a155-104a9018b86e>

...notes...

Y = Number of Weekly Travelers

X = Each Feature

Create a Four Trendlines

<https://support.microsoft.com/en-us/office/add-a-trend-or-moving-average-line-to-a-chart-fa59f86c-5852-4b68-a6d4-901a745842ad>

Evaluate the Correlation

...notes...

Correlation Coefficient

<https://www.thoughtco.com/how-to-calculate-the-correlation-coefficient-3126228>

Files: [fileIddownload.png](https://www.mindmeister.com/generic_files/get_attachment/11548480)

Use "=CORREL(..."

<https://support.microsoft.com/en-us/office/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92>

Model

...develop a mathematical model that captures the essence of the system...

Review Each Trendline

...notes...

Call the Trendline Equation

<https://www.ablebits.com/office-addins-blog/2019/01/16/excel-trendline-types-equations-formulas/>

Intepret the "y = (f) x + e" Equation

<https://nhsexcel.com/forecasting_linear_trendline/>

...more...

Forecasting using Excel’s Linear Trendline

<https://nhsexcel.com/forecasting_linear_trendline/>

Review the R-squared Metric

<https://statisticsbyjim.com/regression/interpret-r-squared-regression/>

Create a Regression Model

...recap...

Regression Fundamentals

<https://docs.google.com/drawings/d/1QovfMtB3tXIXrF5wzvVbEHf-QC2861Cbhxek-SFakso/edit?pli=1>

...steps...

Activate the "Analysis ToolPak" Add-In

<https://support.microsoft.com/en-us/office/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4>

Create a Simple Linear Regression Model

...notes...

Y = Number of Weekly Travelers

X = Price per Week

Evaluate the Simple Linear Regression Model

...notes...

R-Squared > 0.75

<https://statisticsbyjim.com/regression/interpret-r-squared-regression/>

p-Value < 0.05

<https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/#:~:text=Regression%20analysis%20is%20a%20form,correlation%20with%20the%20dependent%20variable.>

Create a Multivariate Linear Regression Model

...notes...

Y = Number of Weekly Riders

X = Price per Week

X = Population

X = Parking Rates

X = Income

Evaluate the Multivariate Linear Regression Model

...notes...

R-Squared > 0.75

<https://statisticsbyjim.com/regression/interpret-r-squared-regression/>

p-Value < 0.05

<https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/#:~:text=Regression%20analysis%20is%20a%20form,correlation%20with%20the%20dependent%20variable.>

Optimise the Regression Model

...regression models must be refined as errors can creep in even if the R2 is good...

Avoiding Collinearity - when the independent variables correlate with each other!

<https://blog.clairvoyantsoft.com/correlation-and-collinearity-how-they-can-make-or-break-a-model-9135fbe6936a?gi=59edf43aa44a>

...notes...

Assessing Collinearity in Excel

<https://www.wallstreetmojo.com/correlation-matrix-in-excel/>

Avoiding Overfitting - when we used all of our data to "train" the model, making too focussed on that data alone!

<https://elitedatascience.com/overfitting-in-machine-learning>

...notes...

Bias-Variance Tradeoff

<https://towardsdatascience.com/bias-variance-and-regularization-in-linear-regression-lasso-ridge-and-elastic-net-8bf81991d0c5>

Files: [fileId0\_lPgJFc6kO2pYGCA7.png](https://www.mindmeister.com/generic_files/get_attachment/11498103), [fileId0\_J3H64-OAEw8Wfa3y.png](https://www.mindmeister.com/generic_files/get_attachment/11498104)

Fixes

Study Correlation between the Xs

Reserve a "Test" Set of Data

[https://cdn.hackernoon.com/hn-images/1\*oU3LAye3LxFcHg0UePmbSA.png](https://cdn.hackernoon.com/hn-images/1%2AoU3LAye3LxFcHg0UePmbSA.png)

...notes...

A Brief Overview of Automatic Machine Learning Solutions

<https://hackernoon.com/a-brief-overview-of-automatic-machine-learning-solutions-automl-2826c7807a2a>

Workflow of a Machine Learning Project

<https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94>

Apply the Regression Models

...we can use the regression models to predict or prescribe...

Predict - define Xs and observe Y

<https://watermark.silverchair.com/tdz011.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAArkwggK1BgkqhkiG9w0BBwagggKmMIICogIBADCCApsGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMjl22Ue_zybLR6YDJAgEQgIICbBqPYNFXBrAfe1YB2-A8LisQccRcXa81aL3RmsSEY7IJ8gdkXC3ku2MpPcY68b6jFICaBkJwmo_r8Av112sgTV_YZrqRsJQFPSjWg8CTYVJPK2LvqTCbBPO83aqkp6MEsJ77nUZ0ulVFWgOFR5BmlqdVSzXjlaKsG_M2sXnWNINDvCdTeSx9JMFS2Lu1lTqlZ0SQOM0R7SH4n4un8GkA8XHH9ZRZl00bwz_nq9Etmw_t_dQmYKEXEAxdJguzk5CCOBWRZUG4SjSclAbsL_RwKK51pcrTrrlm_hSGgYVbHvSx9qdbevyokYM7zVcc6NwZqT1FuujbylevmIcGQHSaOBN-Ur3x38hPHV5zW1p_X7cmh0TAVnXl4Pj_lXFRfHJMEjef-sj6tRjFT6NMbpYZJ97AewizxM-wU8xD4tejHM02wPiVq0uDc9HS03OcOXQXvQ1RmX1-QYNt1swYX46IFm6cRYSrbB0k2iqgPvdhO1UQvrMG0LJB9aYyV_Nwy0Tilp44rx_AM21ya6ZqbkywZLbLVgJNbLIvL8p_1a6wj66MscmMbLlYYPqmREEjYFABja3-Z3EFGu4yzi9NYxHgcR8e5fs7V5f9hAnuoQZiUKucj2YmKeaxLd7Q2FNqDGe620BgkiKs2cuQYrHO27ruKgpFWPEuUe6meaYMkY6hH-CkCyvTQw0xVjuhrgTPaMCnYp5BSoAaZnu-2zly7JspxKiv-KcmqkVvxVhhaRtDe2Zntj3nWubKzaA6BuTZWujDQDOyGPus-cRp0QjMRK2eCa_OlFz8SVxSVac61BomFcwt27zBQfXcxG5fw7oD>

...more...

Linear Regression Analysis in Excel

<https://www.ablebits.com/office-addins-blog/2018/08/01/linear-regression-analysis-excel/>

Prescribe - define desired Y and observe Xs

<https://engineerexcel.com/how-to-use-excel-solver-for-multiple-linear-regression/>

...recap...

Predict an Event or Classify a Case

<https://www2.deloitte.com/us/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html/>

In Power BI

...framing...

Power BI is a Business Intelligence Tool

<https://energycentral.com/sites/default/files/ec-core/nodes/227281/new_value_curve_-_oct_2015_-_final.jpg>

A Data Scientist can use Power BI to Probe Features and Patterns in Big Datasets

<https://powerbi.microsoft.com/en-us/partner-showcase/>

...learning...

Question

...all data science must begin with a question...

Total Sales of Each Sales Person?

Total Returns of Each Sales Person?

Query

...connecting to datasets...

Power BI Data Connections Guide

<https://docs.microsoft.com/en-us/power-bi/connect-data/>

Our Datasets

Files: [fileIdPerson.txt](https://www.mindmeister.com/generic_files/get_attachment/13470332), [fileIdReturns.csv](https://www.mindmeister.com/generic_files/get_attachment/13470333), [fileIdOrders.xlsx](https://www.mindmeister.com/generic_files/get_attachment/13470334)

Wrangle

...cleaning data for use...

Power Query Editor Guide

<https://docs.microsoft.com/en-us/power-query/power-query-ui>

Promote Headers

Model

...form relationships between datasets...

Power BI Data Models Guide

<https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-create-and-manage-relationships>

Person Region > Orders Region

Returns Order ID > Orders Order ID

Enhance

...use DAX to calculate useful metrics...

Power BI DAX Guide

<https://docs.microsoft.com/en-us/dax/dax-function-reference>

Use "=CALCULATE(..." in New Measure

<https://docs.microsoft.com/en-us/dax/calculate-function-dax>

Visualise

...create insightful charts...

Power BI Charts Guide

<https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a>

Create a Chart that Reveals Sales Person Sales Performance

 Problem

...the chart is not delivering the results we expect, we need to diagnose the defect...

What element is not working as it should?

What defects could explain this outcome?

Which defect is the problem in this case?

Which tool might be able to fix this defect?

...doing...

Returning to Features, Patterns, and Models

Question

...notes...

What determines how long a person lives?

Query

...notes...

Our Data

Files: [fileIdgapminder.txt](https://www.mindmeister.com/generic_files/get_attachment/13473709)

Wrangle

...notes...

Address Gaps in Population and Income by Filling Down in Power Query Editor

<https://www.excelcampus.com/powerquery/fill-down-blank-null-cells-power-query/>

Model

...notes...

One Table = The Data Model

Enhance

...notes...

Use "=CALCULATE(..." in New Measure to get the Life Expectancy for a Country and a Year

Visualise

...notes...

Create charts that reveal the relationship between Life Expectancy and other variables (i.e. Features and Patterns)

Employ Slicers to Enable this Search

<https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-slicers>

Interpret

...notes...

What features appear to influence Life Expectancy?

Second Cycle

...this challenge has been faced by leading data scientists...

Hans Rosling

<https://www.youtube.com/watch?v=jbkSRLYSojo>

...more...

Gapminder Visual

<https://s3-eu-west-1.amazonaws.com/static.gapminder.org/GapminderMedia/wp-uploads/20161215212516/countries_health_wealth_2016_v151.pdf>

...more...

 Replicate the Gapminder visualisation in Power BI Desktop...

Files: [fileId3.\_Replicating\_Gapminder.txt](https://www.mindmeister.com/generic_files/get_attachment/12599414)

Visualise+

...notes...

Recreate the Gapminder Visualisation in Power BI

Interpret+

...notes...

How might we capture the relationship we see between Life Expectancy and Income in a Mathematical Model?

...recap...

Forecasting using Excel’s Linear Trendline

<https://nhsexcel.com/forecasting_linear_trendline/>

 Problem

...the line pattern is not straight...

Types of Line

<https://www.statisticsfromatoz.com/uploads/7/3/2/1/73216723/4-curves_orig.png>

...more...

Linearisation

<https://www.myhaikuclass.com/davidyoung.djy/scientificinquiryskills/cms_page/view/21918311>

Files: [fileIdattachment\_65032176.jpg](https://www.mindmeister.com/generic_files/get_attachment/13470425)

Visualise+

...notes...

Format the x-Axis to Log

<https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-customize-x-axis-and-y-axis>

...we can develop our analysis to reflect this...

Query

...notes...

Create a Second Copy of the Data

Files: [fileIdgapminder.txt](https://www.mindmeister.com/generic_files/get_attachment/13473709)

Wrangle

...notes...

Address Gaps in Population and Income by Filling Down in Power Query Editor

<https://www.excelcampus.com/powerquery/fill-down-blank-null-cells-power-query/>

Use Filters to Isolate One Year

<https://support.microsoft.com/en-us/office/filter-data-power-query-b5610630-f5bf-4ba4-9217-a628f9b89353>

Model

...notes...

One Table = The Data Model

Enhance

...notes...

Use "=LOG10(..." in New Column to Create a Logarithmic Income Column

Visualise

...notes...

Y = Life Expectancy

X = Logarithmic Income

Detail = Country

Visualise+

...notes...

Add a Trendline to the Chart

<https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-analytics-pane>

Enhance+

...notes...

Calculate a Correlation Coefficient for Life and Log Income Using Quick Measures and Plot as a Card

<https://exceltown.com/en/tutorials/power-bi/powerbi-com-and-power-bi-desktop/dax-query-language-for-power-bi-and-power-pivot/pearson-correlation-coefficient/>

Interpret

...notes...

What have we shown?

#### ...it is useful to reflect on the differences and complementarities of design thinking and data science approaches...

##### Revealing Systemic Insights

##### Capturing Cause and Effect

### Idea

#### ...the objective of the "Ideation" phase is to craft a response to the insights gathered...

##### Design Thinking

###### ...the design thinking approach focusses on structured brainstorms...

Design Thinking Double Diamond

Files: [fileId1\_4ITg3Y27JfVlnKPPA9pJlA.png](https://www.mindmeister.com/generic_files/get_attachment/12660174)

##### Data Science

###### ...in pursuit of solutions, a data scientist might leverage machine learning...

Machine Learning

...the term "Machine Learning" represents a set of mathematical models that are trained to perform tasks and which improve with experience of that task...

Machine Learning Family

<https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

Files: [fileId1\_8wU0hfUY3UK\_D8Y7tbIyFQ.png](https://www.mindmeister.com/generic_files/get_attachment/11082805)

...more...

Machine Learning Types

Files: [fileId841de56d902740eb5ad461fb8cb9646b.png](https://www.mindmeister.com/generic_files/get_attachment/12015785)

...examples...

Unsupervised Learning

<https://towardsdatascience.com/customer-segmentation-using-k-means-clustering-d33964f238c3>

Reinforcement Learning

<https://towardsdatascience.com/a-beginners-guide-to-reinforcement-learning-with-a-mario-bros-example-fa0e0563aeb7>

Supervised Learning

<https://usedcarpricepredictor.herokuapp.com/>

...recap...

Regression

<https://docs.google.com/drawings/d/1QovfMtB3tXIXrF5wzvVbEHf-QC2861Cbhxek-SFakso/edit?pli=1>

...each category of machine learning methods has a specific set of mathematical tools and processes...

Machine Learning Algorithms

<https://d1jnx9ba8s6j9r.cloudfront.net/blog/wp-content/uploads/2018/03/Types-of-Machine-Learning-Waht-is-Machine-Learning-Edureka-2.png>

Machine Learning Processes

<https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/>

...more...

A Quick Guide to How Machines Learn

<https://www.boozallen.com/content/dam/boozallen_site/sig/pdf/publications/machine-intelligence-quick-guide-to-how-machines-learn.pdf>

...the majority of cases come down to a choice between Unsupervised and Supervised methods...

Supervised vs. Unsupervised

<https://blog.seebo.com/wp-content/uploads/2019/01/6.png>

...more...

Supervised vs. Unsupervised

<https://blog.seebo.com/supervised-vs-unsupervised-machine-learning/>

...the predictive and prescriptive potential of Supervised methods makes it interesting to us when thinking about design...

Predict an Event or Classify a Case

<https://www2.deloitte.com/us/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html/>

Predicting Booking Defaults

<https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=3446&amp;context=sis_research>

Predict a Value

<https://usedcarpricepredictor.herokuapp.com/>

Supervised Learning

...Supervised Learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs...

Supervised Learning Overview

<https://www.tibco.com/sites/tibco/files/media_entity/2020-09/supervised-learning-diagram.svg>

...more...

What is Supervised Learning?

<https://www.tibco.com/reference-center/what-is-supervised-learning>

...it is distinguished by its "Test-Train" process pattern...

Supervised Learning Process Pattern

<https://hackernoon.com/a-brief-overview-of-automatic-machine-learning-solutions-automl-2826c7807a2a>

Files: [fileId1\_oU3LAye3LxFcHg0UePmbSA.png](https://www.mindmeister.com/generic_files/get_attachment/12636385)

...more...

Indicio ML Workflow

<https://indico.io/wp-content/uploads/2015/05/building-ML-models-inhouse.jpg>

Files: [fileIdbuilding-ML-models-inhouse.jpg](https://www.mindmeister.com/generic_files/get_attachment/11493218)

...we can highlight key components of this process workflow...

Sampling

Sampling: collecting data that is representative of the system being studied...

Feature Engineering

Feature Engineering: labelling the data - input and output variables - and exploring which input variables are predictive of the output...

Splitting

Splitting: creating two sets of data - one to train the model and one to test the results - ensuring that the model is not overfitted...

...notes...

Supervised Learning Workflow

<https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94>

Overfitting Overview

[https://miro.medium.com/max/2112/1\*40vADqa\_tlFLiLb4320V7w.png](https://miro.medium.com/max/2112/1%2A40vADqa_tlFLiLb4320V7w.png)

...more detail...

The Case of Overfitting

<https://elitedatascience.com/overfitting-in-machine-learning>

Files: [fileId320px-Overfitting.svg.png](https://www.mindmeister.com/generic_files/get_attachment/11493343)

Training

Training: selecting and executing an appropriate supervised learning algorithm...

...notes...

Machine Learning Model Selection

<https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html>

Files: [fileIdml\_map.png](https://www.mindmeister.com/generic_files/get_attachment/11082909)

Supervised Learning Family

<https://de.mathworks.com/help/stats/machine-learning-in-matlab.html>

Testing

Testing: evaluating the performance of the trained model and assessing its ability to perform its purpose...

...notes...

ML Model Performance Metrics

<https://www.kdnuggets.com/2018/06/right-metric-evaluating-machine-learning-models-2.html>

...with this knowledge we see that machine learning process fits well with the CRISP-DM convention...

CRISP-DM

<https://www.tibco.com/sites/tibco/files/media_entity/2019-03/data_science-diagram.svg>

...more...

What is Data Science?

<https://www.tibco.com/reference-center/what-is-data-science>

ML at Uber

<https://1fykyq3mdn5r21tpna3wkdyi-wpengine.netdna-ssl.com/wp-content/uploads/2018/11/image6.png>

...more...

Uber Michaelangelo

<https://eng.uber.com/scaling-michelangelo/>

###### ...we can experience and develop these concepts in the following scenarios...

In Excel

...we have already completed an ML exercise in Excel...

Q: Can I Predict and Control Commuting Behaviour in a City?

Our Data

Files: [fileIdTransport.xlsx](https://www.mindmeister.com/generic_files/get_attachment/11546020)

What is Missing?

In KNIME

...framing...

Data Science Presents Tech Hurdles

<https://dev.to/minchulkim87/my-data-science-tech-stack-2020-1poa>

Innovative Tools Help

<https://www.tibco.com/resources/analyst-report/gartner-magic-quadrant-data-science-and-machine-learning-platforms>

KNIME is One Such Tool

<https://www.knime.com/>

...more...

KNIME

<https://en.wikipedia.org/wiki/KNIME>

...learning...

Familiarise with KNIME

...steps...

Review Workbench

...notes...

Keep "Node Repository"

Keep "Workflow Editor"

Review a KNIME Workflow

...steps...

Open Workflow

...notes...

Our Workflow

Files: [fileIdSimple\_Statistics.knwf](https://www.mindmeister.com/generic_files/get_attachment/13473673)

Create an EDA Workflow

...steps...

Load Required Tools

...notes...

File > Install KNIME Extensions

Search for "KNIME JavaScript Views (Labs)"

Select > Next > Restart

Locate and Prepare Data

...notes...

Our Data

Files: [fileIdExit.csv](https://www.mindmeister.com/generic_files/get_attachment/13473863)

Create New Workflow Canvas

...notes...

New > New KNIME Workflow > Finish

Select and Configure Nodes

...notes...

File Reader Node

...a node that accesses data in files...

Node Repository > File Reader > Drag to Canvas

File Reader Node > Double Click > Browse > Exit.csv > Review Node Configuration > Apply > Ok...

Data Explorer Node

...a node that prepares an EDA overview...

Node Repository > Data Explorer > Drag to Canvas

Join Nodes

...notes...

Black "Out" Triangle to Black "In" Triangle

Review Traffic Light Indicators

Execute Workflow

...notes...

Right Click Data Explorer Node > Select "Execute and Open Views"

Right Click Data Explorer Node > Select "Interactive View" > Consider Exclusions

Save and Close Workflow

...notes...

File > Save

Translate Gapmider to KNIME

...steps...

Load Required Tools

...notes...

None Required

Locate and Prepare Data

...notes...

Our Data

Files: [fileIdgapminder.txt](https://www.mindmeister.com/generic_files/get_attachment/13473709)

Create New Workflow Canvas

...notes...

New > New KNIME Workflow > Finish

Select and Configure Nodes

...notes...

Join Nodes as Created

...notes...

Black "Out" Triangle to Black "In" Triangle

Review Traffic Light Indicators

File Reader Node

...a node that accesses data in files...

Node Repository > File Reader > Drag to Canvas

File Reader Node > Double Click > Browse > Gapminder.txt > Review Node Configuration > Apply > Ok...

Missing Value Node

...a node that fills gaps in datasets...

Node Repository > Missing Value > Drag to Canvas

Missing Value Node > Double Click > Set to Previous Value > Apply > Ok...

Row Filter Node

...a node that filters data by row values...

Node Repository > Row Filter > Drag to Canvas

Row Filter Node > Double Click > Include Rows by Attribute Value > Set to Year > set to 2015 > Apply > Ok...

Colour Manager Node

...a node that assigns colours to data...

Node Repository > Colour Manager > Drag to Canvas

Colour Manager Node > Double Click > Toggle Colour Range > Apply > Ok...

Colour Manager Node > Right Click > View "Table with Colours"...

Scatter Plot Node

...a node that plots scatter charts...

Node Repository > Scatter Plot > Drag to Canvas

Scatter Plot Node > Double Click > Set X Axis to Income > Set Y Axis to Life > Apply > Ok...

Join Nodes

...notes...

Black "Out" Triangle to Black "In" Triangle

Review Traffic Light Indicators

Execute Workflow

...notes...

Click "Execute All Nodes" in Toolbar

Scatter Plot Node > Right Click > Execute and Open Views...

Select and Configure Nodes+

...notes...

Join Nodes as Created

...notes...

Black "Out" Triangle to Black "In" Triangle

Row Filter > Math Formula > New Scatter Plot

Review Traffic Light Indicators

Math Formula Node

...a node that performs calculations in workflows...

Node Repository > Math Formula > Drag to Canvas

Math Formula Node > Double Click > Set Expression = "log($income$)" > Set Append as "Income\_New"> Apply > Ok...

Scatter Plot Node

...a node that plots scatter charts...

Node Repository > Scatter Plot > Drag to Canvas

Scatter Plot Node > Double Click > Set X Axis to Income\_New > Set Y Axis to Life > Apply > Ok...

Join Nodes+

...notes...

Black "Out" Triangle to Black "In" Triangle

Review Traffic Light Indicators

Execute Workflow+

...notes...

Click "Execute All Nodes" in Toolbar

Scatter Plot Node > Right Click > Execute and Open Views...

Save and Close Workflow

...notes...

File > Save

...doing...

Machine Learning Workflows

...recap...

Supervised Learning Process Pattern

<https://hackernoon.com/a-brief-overview-of-automatic-machine-learning-solutions-automl-2826c7807a2a>

Files: [fileId1\_oU3LAye3LxFcHg0UePmbSA.png](https://www.mindmeister.com/generic_files/get_attachment/12636385)

...more...

Indicio ML Workflow

<https://indico.io/wp-content/uploads/2015/05/building-ML-models-inhouse.jpg>

Files: [fileIdbuilding-ML-models-inhouse.jpg](https://www.mindmeister.com/generic_files/get_attachment/11493218)

Prediction

...workflow components...

Data

...notes...

Our Data

Files: [fileIdCities.csv](https://www.mindmeister.com/generic_files/get_attachment/13473818)

Nodes

...notes...

Input Path

...notes...

File Reader Node

Missing Value Node

Scatter Matrix Node

Partitioning Node

Train Path

...notes...

Linear Regression Learner Node

Test Path

...notes...

Regression Predictor Node

Numeric Scorer Node

Output Path

...notes...

Excel Writer Node

Diagnostic Path

...notes...

Excel Writer Node

Execute

...notes...

Right Click Learner > Review Outputs

Right Click Predictor > Review Outputs

Right Click Scorer > Review Outputs

Iterate

...notes...

Feature Selection

Model Selection

...notes...

Random Forest Learner (Regression)

Random Forest Predictor (Regression)

Numeric Scorer

...reference...

Linear Regression

<https://hub.knime.com/sreekanthnair/spaces/Public/latest/01_Linear_Regression_solution~kvq1INVgpL6zIz7k>

Random Forest Regression

[https://hub.knime.com/raksharawat/spaces/Public/latest/project%2F9.%20Random%20Forest%20Regression~6dkpKR7RShJyNV9d](https://hub.knime.com/raksharawat/spaces/Public/latest/project/9.%20Random%20Forest%20Regression~6dkpKR7RShJyNV9d)

Classification

...workflow components...

Data

...notes...

Our Data

Files: [fileIdExit.csv](https://www.mindmeister.com/generic_files/get_attachment/13473863)

Nodes

...notes...

Input Path

...notes...

File Reader Node

Missing Value Node

Statistics Node

Partitioning Node

Train Path

...notes...

Logistic Regression Learner Node

Test Path

...notes...

Logistic Regression Predictor Node

Scorer Node

Output Path

...notes...

Excel Writer Node

Diagnostic Path

...notes...

Excel Writer Node

Execute

...notes...

Right Click Learner > Review Outputs

Right Click Predictor > Review Outputs

Right Click Scorer > Review Outputs

Iterate

...notes...

Feature Selection

Model Selection

...notes...

Random Forest Learner (Regression)

Random Forest Predictor (Regression)

Numeric Scorer

...reference...

Logistic Regression

[https://hub.knime.com/knime/spaces/Examples/latest/04\_Analytics%2F04\_Classification\_and\_Predictive\_Modelling%2F06\_Logistic\_Regression~SOUGsGodd-2vs8HJ](https://hub.knime.com/knime/spaces/Examples/latest/04_Analytics/04_Classification_and_Predictive_Modelling/06_Logistic_Regression~SOUGsGodd-2vs8HJ)

Predicting Employee Attrition with Machine Learning

<https://www.knime.com/blog/predicting-employee-attrition-with-machine-learning>

In KNIME+

...extension...

How Might We Craft a Predictive Model for the Gapminder Case and Combine KNIME and Power BI Functionality?

#### ...it is useful to reflect on the differences and complementarities of design thinking and data science approaches...

##### Exploratory Nature of Design Thinking

##### Constrained "Predict and Prescribe" Intent of Machine Learning

### Action

#### ...the objective of the "Action" phase is to translate and package our ideas in a way that generates meaningful impact...

##### Design Thinking

###### ...in the world of "Design Thinking" we can highlight two interesting themes in the "Action" phase...

Prototyping as Learning

<https://www.interaction-design.org/literature/article/stage-4-in-the-design-thinking-process-prototype>

Files: [fileId27e5f3617d49e47ae93991a8d28a098d.jpg](https://www.mindmeister.com/generic_files/get_attachment/11120359)

...examples...

Type: App Wireframes

<https://www.usability.gov/how-to-and-tools/methods/wireframing.html>

Files: [fileIdWhy-Wireframing-of-Your-Mobile-App-Is-So-Important-05.webp](https://www.mindmeister.com/generic_files/get_attachment/11120498)

...tool...

Marvel Apps

<https://marvelapp.com/signin?next=/project/4013333>

Type: Solution Canvasses

<https://www.strategyzer.com/canvas/business-model-canvas>

Type: Storyboards

<https://blog.prototypr.io/all-that-you-need-to-try-google-material-design-8395df3902c>

Files: [fileId1\_BUnG41ggf4pYXE306Ygktg.png](https://www.mindmeister.com/generic_files/get_attachment/11548893)

...more...

Customer Labs

<https://www.customerlabs.co/blog/the-power-of-prototype-in-design-thinking/>

Career Foundary

<https://careerfoundry.com/en/blog/ux-design/design-thinking-stage-four-prototyping/>

Law by Design

<https://www.lawbydesign.co/design-process/>

Feasibility-Viability-Desirability

<https://designthinking.ideo.com/>

...more...

Desirability, Feasibility, Viability: The Sweet Spot for Innovation

<https://medium.com/innovation-sweet-spot/desirability-feasibility-viability-the-sweet-spot-for-innovation-d7946de2183c>

###### ...one tool stands out as a good point of reference...

Service Blueprint

<https://media.nngroup.com/media/editor/2017/08/22/nng-service-blueprint-example.png>

...more...

Service Blueprint Template

<https://media.nngroup.com/media/editor/2017/08/22/nng-service-blueprint-101.png>

Service Blueprints Overview

<https://www.nngroup.com/articles/service-blueprints-definition/>

##### Data Science

###### ...we can focus attention on two approaches from the data sciences that can add value to the "Action" phase...

Logic Models

...task...

Craft a Theory of Change, Logic Model, and Indicator Set for your idea...

...help...

Theory of Change Worksheet

[https://design-kit-production.s3-us-west-1.amazonaws.com/Design+Kit+Method+Worksheets/DK\_theoryofchange\_activityguide.pdf](https://design-kit-production.s3-us-west-1.amazonaws.com/Design%2BKit%2BMethod%2BWorksheets/DK_theoryofchange_activityguide.pdf)

Logic Model Template

Files: [fileIdAnnotation\_2020-04-18\_161201.png](https://www.mindmeister.com/generic_files/get_attachment/11120457)

Indicators Guide

[https://design-kit-production.s3-us-west-1.amazonaws.com/Design+Kit+Method+Worksheets/DesignKit\_indicatormadlibs\_worksheet.pdf](https://design-kit-production.s3-us-west-1.amazonaws.com/Design%2BKit%2BMethod%2BWorksheets/DesignKit_indicatormadlibs_worksheet.pdf)

...there is an area of work that focusses on instrumenting a process - collecting data and driving optimisation...

Service Instrumentation

<https://www.mckinsey.com/business-functions/operations/our-insights/four-ways-to-shape-customer-experience-measurement-for-impact>

Indicators

<https://www.designkit.org/methods/define-your-indicators>

...more...

Indicators Guide

[https://design-kit-production.s3-us-west-1.amazonaws.com/Design+Kit+Method+Worksheets/DesignKit\_indicatormadlibs\_worksheet.pdf](https://design-kit-production.s3-us-west-1.amazonaws.com/Design%2BKit%2BMethod%2BWorksheets/DesignKit_indicatormadlibs_worksheet.pdf)

...this links to the economists' "Logic Chain" concept...

A Logic Model is a succinct and visual narrative that summarizes the final set of activities in your solution, and the way that each of these will lead to change over time...

Climate Change Adaptation

<https://www.ukcip.org.uk/wizard/adaptme-toolkit/fundamentals/logic-and-assumptions/>

Files: [fileIdAdaptME-adaptation-logframe-large.png](https://www.mindmeister.com/generic_files/get_attachment/11120400)

...more...

Logic Model Template

Files: [fileIdAnnotation\_2020-04-18\_161201.png](https://www.mindmeister.com/generic_files/get_attachment/11120457)

Logic Model

<https://www.designkit.org/methods/create-a-logic-model>

Theory of Change

<https://www.designkit.org/methods/explore-your-theory-of-change>

Theory of Change Worksheet

[https://design-kit-production.s3-us-west-1.amazonaws.com/Design+Kit+Method+Worksheets/DK\_theoryofchange\_activityguide.pdf](https://design-kit-production.s3-us-west-1.amazonaws.com/Design%2BKit%2BMethod%2BWorksheets/DK_theoryofchange_activityguide.pdf)

The Magenta Book

Files: [fileIdHMT\_Magenta\_Book.pdf](https://www.mindmeister.com/generic_files/get_attachment/11120390)

...Logic Chains are geared for instrumentation...

4 x Es and VfM

<https://beamexchange.org/guidance/monitoring-overview/assessing-value-money/4e-approach-vfm/>

Files: [fileIdmonitoring-4e.svg](https://www.mindmeister.com/generic_files/get_attachment/11120536)

Confirmatory Experimentation

...a large portion of data science concerns the question: "Did it Work?"...

Analytics Value Curve

<https://energycentral.com/sites/default/files/ec-core/nodes/227281/new_value_curve_-_oct_2015_-_final.jpg>

...like the rest of the data sciences, getting good outcomes from experimentation requires strong process discipline...

Experimental Design

...experimentation is about tackling "Probability", "Significance", and "Inference"...

Experimental Pattern

<https://bolt.mph.ufl.edu/6050-6052/unit-1/>

Files: [fileIdimages-mod1-big-picture-eda.gif](https://www.mindmeister.com/generic_files/get_attachment/11548581)

...example...

A-B Testing

<https://unbounce.com/conversion-rate-optimization/small-changes-increasing-conversion-rates/>

Files: [fileIdconversion-social-proof.png](https://www.mindmeister.com/generic_files/get_attachment/13477545)

...more...

The Big Picture

<https://bolt.mph.ufl.edu/6050-6052/preliminaries/the-big-picture/>

...we need a grasp of some fundamentals of the "Experimental Pattern" in order to shape these experiments...

Experimental Design Pattern

<https://docs.google.com/presentation/d/18caI4xY7d3qC1fa5JdM9nSbiEUvCZpyH7nMFkwDFIgo/edit?usp=sharing>

...more details...

Randomised Control Trials

Files: [fileIdIllustration-of-a-randomised-controlled-trial-RCT-to-test-a-new-back-to-work.png](https://www.mindmeister.com/generic_files/get_attachment/11123906)

Isolating Variables (Blocking)

<https://introductorystats.wordpress.com/2011/03/09/design-of-experiments/>

Experimental Design

<https://www.scribbr.com/methodology/hypothesis-testing/>

Type 1 and 2 Errors

<https://www.simplypsychology.org/type_I_and_type_II_errors.html>

Files: [fileIdtype-1-and-2-errors.webp](https://www.mindmeister.com/generic_files/get_attachment/11548587)

Statistical Tests

<https://towardsdatascience.com/statistical-tests-when-to-use-which-704557554740>

Representativeness

<https://www.qualtrics.com/blog/calculating-sample-size/>

Hypotheses

<https://www.scribbr.com/research-process/hypotheses/>

Significance

<https://www.dummies.com/education/math/statistics/what-a-p-value-tells-you-about-statistical-data/>

Files: [fileIddownload\_\_1\_.jpg](https://www.mindmeister.com/generic_files/get_attachment/11548588)

...task...

Shape an experimental design to establish whether an MRT nudge campaign has an impact on congestion...

...help...

Research Question

Measured Variable

Sample Size and Profile

Block Structure

Treatment and Protocol

Statistical Test

...background...

SG MRT Nudges

Files: [fileIddownload\_\_2\_.jpg](https://www.mindmeister.com/generic_files/get_attachment/11548624)

Nudge Theory

<https://www.businessballs.com/improving-workplace-performance/nudge-theory/>

Files: [fileIdNudge-Theory.jpg](https://www.mindmeister.com/generic_files/get_attachment/11548619)

Transport System Nudges

<https://datasmart.ash.harvard.edu/news/article/nudging-sustainable-mobility>

...we need some capability running statistical tests if we are going to be successful...

In Excel

...we can get a handle on the tests using Excel scenarios...

Statistical Scenarios

Files: [fileIdDiagnostic\_Analytics.xlsx](https://www.mindmeister.com/generic_files/get_attachment/13477733)

In KNIME

...framing...

Statistical Tests

<http://abacus.bates.edu/~ganderso/biology/resources/stats_flow_chart_v2014.pdf>

Significance and Inference

<https://phhp-faculty-cantrell.sites.medinfo.ufl.edu/files/2012/07/images-mod1-big-picture-eda.gif>

...learning...

T-Tests, ANOVA, and Chi

...steps...

Data Sample

...notes...

Our Data

Files: [fileIdExams.csv](https://www.mindmeister.com/generic_files/get_attachment/13477714)

New Workflow

File Reader Node

Data Explorer Node

Experimental Design

...notes...

Question

Metric

Treatment

Sample

Blocks

Test

Test and Evaluate

...notes...

Independent Groups T-Test Node

One Way ANOVA Node

Crosstab Node

...doing...

Perform an A-B Test

...notes...

Our Data

Files: [fileIdAB.csv](https://www.mindmeister.com/generic_files/get_attachment/13477725)

File Reader Node

Data Explorer Node

Unpivoting Node

Independent Groups T-Test Node