# 

Assignment 2:

[www.pilatesology.com](http://www.pilatesology.com)

user performance in different regions



**Abstract:**

This data-scientific activity analyses the correlation between recognized and opportunity factors, which can lead to customer transformation. It also describes user behaviour in distinct areas.

Presentation link:

<https://youtu.be/FQlCq5BZ8DM>

Lecturer: Dr.Pamela Spink

Student Name: Maryam Mahmoodi

Student ID: 30155843

[Table](#_Toc15834694) of Contents

[1. Project report: 3](#_Toc15834695)

[2. Business Model: 3](#_Toc15834696)

[2.1. Life cycles and value chains: 5](#_Toc15834697)

[2.2. Analytical level: 6](#_Toc15834698)

[Descriptive analytics 6](#_Toc15834699)

[Prescriptive analytics 6](#_Toc15834700)

[2.3. Application area: 7](#_Toc15834701)

[Product and marketing team: 7](#_Toc15834702)

[Business analyst: 7](#_Toc15834703)

[Data science team: 7](#_Toc15834704)

[Engineering: 7](#_Toc15834705)

[2.4. Data management: 7](#_Toc15834706)

[2.4.1 Full lifecycle actions: 8](#_Toc15834707)

[2.4.2 Sequential actions: 9](#_Toc15834708)

[2.4.3 Occasional actions: 10](#_Toc15834709)

[3. Characterising data and data processing 11](#_Toc15834710)

[3.1 Volume: 11](#_Toc15834711)

[3.2 Velocity: 11](#_Toc15834712)

[3.3 Variety: 12](#_Toc15834713)

[3.4 Veracity 13](#_Toc15834714)

[4. Resources 13](#_Toc15834715)

[5. Data analysis 14](#_Toc15834716)

[4. References: 15](#_Toc15834717)

# Project report:

Pilates is a unique exercise focusing on controlled movements and flexure of the body that strengthen and improve muscle and joints. In this project we will focus on an online company called *Pilatesology* that provides its users with a compilation of exercises that they can access from the comfort of their own home as well as a resource for trainers to manage their customers calendar and track their achievements. This project endeavours to explore the business model that will give the online company its best chance of success with special consideration in business structure, finance and marketing.

A study conducted by IBISWORLD shows the factors for an ideal Pilates studio is ease in accessibility (1), equipment (2) and a thorough understanding in how to use it (3).

This report will discuss the variables that influence conversion rates from users in trial period to paying customers and how to segment them according to the Pilatesology audience to create an individualised approach for the program. In order to do this we will consider their location, age and generic online behaviour to create a database that will inform the following:

* Dividing clients base on their online behaviour
* Developing marketing strategies
* Constructing a landing page
* Increasing shopping bag to payment rate

However, to be able to accomplish the above, we first need to address two unknowns.

* Do recognized factors relate to transformation conduct?
* Does behaviour differ between places

# Business Model:

This diagram depicts the decision process that aid the data analysis.



Table 1 is showing the components of various categorises from figure 1 consist of company’s variables, decisions and outcomes.it shows us what the customers behavioural patterns, what they like and use and therefore guides how the company can improve their experience. Consequentially, if the customers are engaged then the business profit margins are accordingly increased, and both sides are satisfied. Therefore, by analysing the data from *Pilatesology* we are able to make informed decisions that will best serve the company.

|  |  |  |  |
| --- | --- | --- | --- |
| **Chance variable** | **Known variable** | **Decisions points** | **Objectives**  **(expected outcome)** |
| Behavior | Current rate | New segment factors | Higher rate |
| Segmentation | Current regional segmentation | Future product | N/A |
| Further product | Available trial users | Website page | N/A |
| N/A | Current product | Trial period | N/A |

Table 1: diagram components

## 2.1. Life cycles and value chains:

The standard value chain may be considered to help us diagnose the possible problems within *Pilatesology* and how we may solve them. The value chain is a full range of activities that visualize analysis the online *Pilatesology* activities and also help us to understand better about the problems and how could to develop and solve the draft model below by add and subsequence the values to the chain to generate the best profit for the business. The online *Pilatesology* is not a big organization compare to a big organization like Google, Facebook, etc. Figure 2 is customized for online *Pilatesology* Data Sciences Project and it is a draft model that gives a visual analysis of the accumulated effects of different data systems and how improving different points of the chain can give a better end result from each time.



Figure 2 Standard Value Chain for Online Pilatesology Data Science Project

There are two stages to the value chain in which will we will There are two phases in the value chain we will explore in *pilatesology*. On the first phase, the main' Data Provide performers ' represents information collected. Team experts handle feedback processing the information. Then the information is designed and controlled regularly. On the second phase, information needs to be formatted and wrangled to pass the outcome for evaluation and transform for display as a visualization to provide the proprietor of the business.

After the information has gone all enhancement phase, the ultimate phase will be reached, and at this point, we can take the information to the intervention to the greatest advantage for the business. It should be noted that each phase is the enhancement phase from the earlier phase and repetition continually until peak efficiency is achieved.



Figure 3: Architecture Taxonomy for online-pilatesology

## 2.2. Analytical level:

Data analytics is the method of information manipulation and analysis to comprehend and uncover ideas that can affect company choices. According to the value chain, data analytics for the online *pilatesology* business has two goals: initially, to explore the behaviour of present clients and, secondly, to provide highest results for potential decisions for this business. The data which has been collected can be organised by the following groups:

### Descriptive analytics

The first aim for this study is, to get more information about client’s behaviour in online *pilatesology* from different regions then navigate the program. On the first two columns on the left of the table 3 shows that the variables have two segments: chance variables and Known variables.

### Prescriptive analytics

Using the data provided, the marketing team decides:

* What are the features engaging the customers the most and what components are not being utilised?
* What helps in ease of use for their customers?
* Which features to include in the trial program?
* Increase the visual appealability of the application?
* How can they reach potential customers?

These all aid for higher revenue, increased customer satisfaction and an overall more successful business.

## 2.3. Application area:

It is required that the different areas within the *Pilatesology* work as a team for the best result possible. The different areas include:

### Product and marketing team:

They are in charge of the flow of data. Within *Pilatesology* they will

* + Fully understand the aims and objectives of the program
  + Identify *Pilatesology* strengths and weaknesses
  + Regulate communication within the different teams and are all informed

### Business analyst:

They are in the grey area between business area and engineering. They are able to aid both areas with whatever they may need

### Data science team:

They are the driving force to convert data into a useable resource. They are responsible for

* + Wrangling the data with the aid of the engineering team
  + Categorise the data into their different characteristics
  + Write up reports for the rest of the teams to use, and to also present to stakeholders to demonstrate the high value of *Pilatesology*
  + With the knowledge of the data, recommend what other steps may be taken to improve the business

### Engineering:

This department regulate the storage and maintenance of the data on the system. They present the data in a manner in which everyone in the team are able to use ( ICPSR, 2017).

* + Identify outliers within the data or chance variables
  + Make the data easy to find based on how it is categorised through available files
  + Format the data with the correct titles and documentations for further analysis
  + Store and document information in a known database format and keep them for last longer

## 2.4. Data management:

1. Online *pilatesology* must fulfil the laws of the Australian government and wellness organization. Because consumers register their personal health information, data is considered private and must be respected. In other words, because the company is engaged in health and wellness, data is considered personal and must be handled securely and legally.

Base on management system there are three levels of operation needed for data management of the DCC curation lifecycle: (Digital Curation Centre University of Edinburgh).

### Full lifecycle actions:

Figure 4, shows all the elements in the lifecycle actions.

Figure 4: lifecycle actions

#### **Description and Representation Information**

To guarantee an adequate long-term description and control, assign administrative, descriptive, technical, structural and conservation metadata using suitable norms. Collect and provide representation data needed for both the digital content and the related metadata to comprehend and make.

#### **Preservation Planning**

A scheme based on solution design is provided by the data architect.

#### **Community Watch and Participation**

Keep an eye on relevant community operations and help develop the common norms, instruments and appropriate software like GitHub.

#### **Curate and Preserve**

data architect and collaboration tools measures scheduled for the entire curation cycle and management.

### Sequential actions:

Figure 5 shows every step in the sequential actions.

Figure 5: Sequential actions

#### **Conceptualise**

Design and schedule information creation, including method and storage choices.

#### **Create or Receive**

It may also add administrative, descriptive, organizational and technical metadata, preservation metadata at inception. Receive information from information makers, other databases, repositories or data centres based on recorded compilation policies and could have appropriate metadata.

#### **Appraise and Select**

Online *pilatesology* must respect Australian government and health agency regulations. Because users record their personal health information, information is regarded private and must be regulated with respect.

#### **Ingest**

Data is recorded and transmitted via secured channels via the mobile app and internet portal and stored in a secure company data centre.

#### **Preservation Action**

Data governance is the exercise of identifying norms process technology that the organization will depend on to handle its information. A thorough data governance strategy meets regulatory requirement insurance business continuity and empowers the collection and recovery of all company information. ([Access Sciences](https://www.youtube.com/channel/UCdMswzbNAmz8j9uhIyNTgLA), 2017)

To ensure data security and governance, Online *pilatesology* passed standard ISO process and captured certificates. Company procedures are audited and preserved by norms.

#### **Store**

Two factor authentication used for access, based on datacentre audited by ISO.

#### **Access Use and Reuse**

Ensure data engineers have day-to-day access to data for designated by secured identities.

#### **Transform**

Create visualisation data based on stakeholder requirements by data scientist.

### Occasional actions:

Figure 5: Occasional actions

#### **Dispose**

**Some data has been selected for long term storage or transfer to another archive, repository or data centre but in some cases, data should be disposed, according to legal policies.**

#### **Reappraise and Migrate**

Engineering group reviews data failing validation procedures for further assessment and re-selection once a year. Migrate data to a separate format. This can be accomplished according to a memory environment or to ensure immunity from the obsolescence of hardware or software.

# Characterising data and data processing

Big Data includes extensive datasets in terms of amount, variety, velocity and/or variation, needing a scalable architecture for efficient processing, processing, and evaluation.

Data science is the empirical synthesis of information actionable through the entire lifecycle. Data is arranged depending on its distinct characteristics to enhance functionality and comfort of use. Therefore, the previous information features are handled as described below:

Figure 6: Characterising data and data processing

## 3.1 Volume:

By introducing green IT technologies and cloud computing, companies have rescued huge quantities of initial spending on constructing technical hardware and software infrastructure. At the same time, most companies are experiencing a significant change in their business in a few early years, leading in bankruptcy and leaving business or significant growth. Either situation needs a notable decrease or cost because they need to manage their IT infrastructure.

In the industry, there are large numbers of cloud-based alternatives that help various applications and deliver exciting business scalability and information consecutively.

Recent advancements within I.T. has enabled newly developing businesses to save money in the initial set up of the company through computing systems such as Cloud. These systems are beneficial to businesses such as *Pilatesology* because they provide potential for growth of both the business, but also the data sample size. Currently, *Pilatesology* is hosted by the Amazon server which gives the business model access to their wide resources, a platform which they can continuously grow and develop with the reassurance that their data is protected.

## 3.2 Velocity:

Velocity is how quickly data arrives. Amazon Web Services (AWS) is a secure cloud services system offering computing energy, database storage, content delivery and other characteristics to help businesses grow and scale. It promotes traditional data assessment on streaming data excerpt, transform, and store (ETL). The data catalogue is automatically created as a permanent metadata store for all data resources, rendering it searchable and quarriable in one view.

AWS Lambda function is used to retrieve data and metadata from multiple file formats. It's a computing software that can operate software without operating computers. AWS Lambda executes the *pilatesology* software only when needed and scales from a few requirements weekly to thousands per second. Otherwise, various buckets can reuse the same Lambda function.

Online *pilatesology* produces data through: Mobile app, Web clicks, Application logs, Enterprise service, Monitoring services, Customer queries and support.

## 3.3 Variety:

Variety concerns organized and unstructured data sets that collected data from tools like spreadsheets and databases. It comes as messages, images, images, surveillance facilities, PDFs, video, etc. This variety of unstructured data creates a memory, mining, and analytical problems.

Figure 7 shows, few Python code lines written in PyCharm to define metadata handling of events to manage data variety. AWS Lambda absorbs these rows and gathers S3 template metadata. Bucket and top item provide S3 metadata. When needed, we can check folder designations. Using get item, we can take the bin expansion and the main file name to the element we're searching for.

def handler(program, context):  
 records = program['Records']  
   
 output\_records = list()  
 for record in records:  
 bucket = record['s3']['bucket']['name']  
 key = record['s3']['object']['key']  
 response = s3.head\_object(Bucket = bucket, Key = key)  
 filename = os.path.basenmae(Key)  
 extension = os.path.splitext(Key)[1][1:]  
 content = s3.get\_object(Bucket = bucket, Key = key)  
 content = json. loads(content['Body'].read())['text']

Figure 7 metadata sample coded using Python

<https://www.youtube.com/watch?v=bC7aJN5L6po&t=1049s>

## 3.4 Veracity

Since data analysis for decision-making is incorporated as part of daily online *pilatesology* activities, veracity is deeply critical and regularly tested. Many training batches screen data models. The company's Trial and Error team run A / B testing as part of daily activities. They acknowledge any discrepancy between evaluation outcomes and original purpose. Other references like interview study findings also verify results.

# Resources

The emergence of open data has significantly improved the data science environment. At Online *pilatesology*, information scientists have much easier links to information and more transparent information. Open data has allowed data scientists within companies such as Pilatesology easier access to their data.

* Coding software and language:

On Amazon Web Services, the company's development environment offers open-source programming on a cloud-based alternative. Data scientist utilizes Python language to perform information cleaning tasks, as well as data analysis and visualization.

AWS identity management enables secure access to encryption and NoSQL-accessible information. As previously mentioned, *Pilatesology* is hosted by Amazon Web Services which allows its programming on a cloud-based solution. The ‘Python’ technical language is utilised to run the data and format it for data evaluation which is available via the NoSQL through Amazon Web Services.

* GitHub

GitHub's online web-based solution provides all team members real-time access to all codes and data resource. GitHub organizes all distributed version control and source code management. The catalogue of API characteristics and specs is kept and accessible on request. Also, it organizes all the distributed version control and source code management.

* Confluence

Confluence provides the ability to migrate from paper-based web creation to wiki-style, web-based content leadership. Confluence is used as a group location to monitor all documents, especially changing requirements. With Confluence, end-users can create cooperative pages, blogs and exchange information. All new applications for confluence architecture layout, safety assessments, sample instances and outcomes etc. are placed in distinct folders.

* Jira

In the lifecycle of company demands, we use Jira to retrieve, maintain and monitor specifications. Data scientists are separate cases, using Jira as a collaborative tool and truth medium to search for answers or communicate with peers to promote projects.

Big data has emerged in the latest years to define multiple terabyte information volumes.

# Data analysis

Based on Figure 1, It is required to use the knowledge of known and unknown variables to guide business decisions to fulfil the already outlined aims of the business. This models shows different choices need to be taken and factors need to be specified. The model is used to predict if the change in these certain variables correlate to the desirable result. The system reflects the series of recognized and unfamiliar factors leading to choices and required results. In other phrases, the model intends to dismiss the hypothesis if recognized and unfamiliar factors correlate with result modifications.

Figure 8 Correlation between variables

Take for example whether the different demographics within *Pilatesology* will respond to the same application by different behavioural patterns or not. To establish an answer, a regression test is run by the data scientist between the known variables and the conversion rate from each demographic. Within the test, controlled variables are used to ensure accuracy. Once this is completed, the data scientist conducts a student’s t-test to compare the behavioural patterns between the different demographics. This t-test will show the different conversion rates between the demographics to see which one is higher and which one is lower.

# References:

Access Sciences(2017, August 7).*www.youtube.com.* Retrieved from **Data Governance | Components of an Effective Data Governance Program:**

https://www.youtube.com/watch?v=t4IOS5csv40

AWS Online Tech Talks. (2017, July 13). *Building Metadata Catalogs for Data Lakes with Elasticsearch Service - 2017 AWS Online Tech Talks.* Retrieved from AWS Online Tech Talks - Youtube channel: https://www.youtube.com/watch?v=bC7aJN5L6po&t=1049s

Digital Curation Centre University of Edinburgh. (n.d.). *DCC Curation Lifecycle Model.* Retrieved from DCC: http://www.dcc.ac.uk/resources/curation-lifecycle-model

ICPSR. (2017, March 6). *www.youtube.com.* Retrieved from ICPSR Summer Program in Quantitative Methods of Social Research : https://www.youtube.com/watch?v=ZEkqF8cL2qQ