Applied Data Analysis Techniques for Businesses

Assignment Report

Techniques/Methods for Analysing Data for Landscape Design Application

Student Name: Vivian Omo-Ojugo
Student Code: S4100903
Module Title: CT7203 Applied Data Analysis Techniques for Businesses
Module Tutor: Professor Shujun Zhang
School of Computing and Technology
University of Gloucestershire
17/05/2021

1 Introduction

Data Science, spans across a wide range of disciplines and when we say data analysis for business, we are not simply referring to the day to day buying and selling which is popularly known. Data analysis techniques for businesses, simply gives an in-depth knowledge of how data analysis can be applied across different disciplines.

This report would focus on select techniques to analyse data for landscape design, the application of these techniques, and the software that can be used for these techniques.

The case study used in this report is McCarthy Taylor Systems Ltd. An independent software house which specialises in 3D terrain modelling software.

2 Business Case

McCarthy Taylor systems limited is an independent software company that produces software for civil engineering and surveying fields. It is located in Aerial view, Acorn house, Shab Hill, Gloucestershire, GL4 8JX. United Kingdom. The company specialises in 3D terrain modelling software, which is used in a large range of companies and applications such as land surveys, geotechnical engineering, construction, hydrographic design and surveying, mineral extraction, landscape architecture, waste management amongst others.

The company started as a partnership in 1985 and became an independent software company in 1995. It was birthed and driven by a need to offer a type of software, which is affordable coupled with a high level of efficiency, to its users which are the Surveyor, Engineer and Designer.

McCarthy Taylor systems takes pride in the high standard of its software called LSS (Land System Surveying) with a backup service to follow suit. This product is a windows-based software which can import a wide range of 3D data. It is a complete land survey terrain modelling, volume design and visualisation package which has about 2500 Engineering and survey experts using it in the United Kingdom alone. The software is highly rated and relied upon by quite several organisations. It is the system of choice for modelling and design by Engineering and survey professionals.

McCarthy Taylor's existing software product (LSS) was built around a 32bit FORTRAN language and compiler. After some time, development on the system,

stopped and this did not mean that the product was outdated as there was a continuous influx of customers into the business because of the product’s innovative, yet straightforward interface and the quality of personal support. However, the existing software, at the time was not future-proofed. It could not take advantage of 64bit computing platforms or utilise multi-threading or multiple CPUs. Developments in the field of terrestrial laser scanners meant that the sheer quantity of data many users desired to import into McCarthy Taylor's software for modelling and analysis, far outstripped the capacity of the existing system. To move the product on, they needed new and transformational development of the existing software, to provide new functional capabilities for manipulating point cloud data. The company indeed had to innovate and not just copy competitors. They had to come up with a semi-automated way of analysing point clouds to extract objects which can then be used to generate 3D line drawings and quantities, such as topographical surveys, building internals or building elevations. With a great deal of mature software available in the market, most of which targeted industrial installations such as oil refineries. There was however an absence of effective software in the company's sector/s, suitable to work with terrain data. McCarthy Taylor however required an injection of skills, expertise, and experience, embedded through the KTP model, to achieve this vision. This development was to see to the long-term growth of the company and provide a UK-sourced solution for hundreds of customers in the UK, many of whom were at the time, reliant on sub-standard imported software that were not affordable or suitable for the intended purpose.

The company's competitive advantage over the last 29 years has been their innovative approach to problem solving. They have an extensive and extremely loyal customer base, many of whom have been with the company since its incorporation. which is a contributing factor to the company’s success. It is these customers, that are the driving force for innovation and continuous improvement. Initially at the start of the company, there were certain features that could not be identified with the software, but with time these have become available. One of such features is the point cloud solution which integrates seamlessly with the software. This addition has once again placed the company ahead of its competitors in areas that would make its clients seek additional services from competing companies.

The three LSS product levels, are LSS Solo, LSS Vista and LSS Elite. They are described according to their product names with Vista and Elite, having additional point cloud modules.

3 Discussion of data analysis techniques

3.1 Requirement Analysis: BPM

Requirement analysis in data analysis, is a technique that involves clearly setting out the user expectations of a project to be carried out. The conditions to be met for the new project have to be determined and while at this, one has to consider the possibility of contradictory views amongst team members.

This technique is highly determinant, in the success or failure of a project. Reasons for investigating and measures to be used for analysis have to be understood.

Requirements should be documented, actionable, measurable, testable, traceable, related to identified project needs and should be sufficiently defined for system design. The process requires joint efforts that combine hardware, software, people management skills, and human factors engineering expertise.

Some important procedures involved in this technique are:

- Identification of client needs.
- System evaluation for feasibility.
- Conduct economic and technical analysis.
- Allocate functions to system elements.
- Establish schedule and limitations.
- Create system definitions.

Business Process Management (BPM)

BPM is defined as “supporting business processes using methods, techniques and software to design, enact, control and analyse operational processes involving humans, organizations, applications, documents and other sources of information. (Ko, 2009)

BPM involve different steps and while some iterate five steps, other experts iterate six steps. Going by five steps, they are (designing, modelling, executing, monitoring, optimization)

Irrespective of the differences, BPM lifecycle generally consist of:

- Design the business process as it should ideally exist and analyse the process as it currently exists and what is needed to improve it;
- Model -- or consider -- how the business process operates in different scenarios.
- Implement -- or execute -- improvement solutions, including standardization and process automation.
- Monitor improvements; and
- Continue to optimize the business process.

Importance of BPM

BPM allows organizational leaders to understand the various processes that happen within their organizations, analyse them from end to end and improve them on an ongoing basis. This activity allows organizational leaders to optimize end-to-end business processes and not simply improve individual tasks, thereby, giving organizational leaders the ability to have a greater impact on outcomes.

Correctly implemented BPM minimises errors and waste, improves time conservation and generally produces better products and services.

BPM is a continuous task therefore company heads, see to the constant continuous management of the business processes therefore, they are committed to finding new ways to improve business processes as the market is faced with new burdens and trends coupled with the fact that new technologies are more efficient in terms of job automation within the overall procedure.

3.2 Collection: 3D Laser Scanners

Data collection technique is the process of collecting, measuring, and analysing accurate insights for research using known acceptable methods. This technique is a highly important step for data analysis. Different areas of study have different methods of data collection with a consideration for what is required. It is important that collected data is reliable to enable statistical analysis for data driven decisions.

Data collection methods can either be:

- Primary or
- Secondary

Primary data collection methods are raw data collected for the first time. While Secondary data collection method is a collection of already published data from either books, online sources, or other means.

One of the methods for data collection in landscape design is the use of 3D laser scanners. An example 3D laser scanner is the EinScan Pro 2x. This scanner is one of the best multifunctional and high-resolution 3D scanners available for small objects. It is equipped with the latest development in data capture hardware in the market with an exceptionally high level of accuracy. Also, it generates high resolution 3D data.

Laser scanning is a popular land surveying method that can accurately measure and collect data from objects, surfaces, buildings, and landscapes. Laser scanners collect information in the form of point cloud data, which consists of millions of 3D coordinates (XYZ coordinates).

In recent times, simple scanners have been replaced with LiDAR systems — high-quality scanners that use monochromatic intelligible light beams to collect object related data.

Modern laser scanners can collect detailed point clouds, and with point cloud processing software, these datasets can create digital 3D models of the scanned environment.

Laser scanners work by throwing out light waves that bounce off of surfaces and then reflects back to the sensor which in turn calculates the distance by measuring the time taken for the light beam to complete its journey. The distance measured is

then used to calculate a coordinate for the tiny section of the surface hit by the laser beam. All of this happens in just seconds, and during a single scan, a laser scanner will collect millions of 3D coordinates.

Some issues associated with laser scanners, are the production of irregular and unordered points. Also, scanners cannot differentiate between noise and points hence noise filtration, is needed.

3.3 Technique 3: Data Cleaning: Outlier removal

Data cleaning is one of the most important steps in data analysis. It involves refining data by identifying incomplete and useless parts of a dataset. Here, irrelevant data is tweaked, or removed until a certain acceptable standard is attained. The cleaning should be done with a view to achieving completeness, validity, accuracy, consistency, and uniformity. failure to clean data accurately, would cause issues to arise during the data analysis phase. This step is critical, as improves the results generated from the data analysis. The process can either be automated or done manually. For the purpose of this report, we will focus on removal of outliers.
This is one key step for optimal performance of your model. Outliers are features that are conspicuously different from the other observations in a dataset.

Data gotten from scanned devices, result in a set of 3D points, called point clouds. These data produced, often contain noise such as flying birds, road furniture, foliage, ghosts and other unnecessary features, it is important that a point cloud, is cleaned, to produce better models and to make it more appropriate for use, by the design team and the employer.

Octree Algorithm

Octree is used to filter all the noise from point cloud data without deleting useful feature point. It first visualises points and then the data is partitioned thereby making it easier to process the data in each node. It is a tree data structure for three-dimensional space partitioning. Octree recursively subdivides the node into eight octants and again divides into eight nodes and this tree structure carries on until a certain number of points are reached in a node called leaf node. Data partitioning by octree, makes it a lot easier to process the information in that node.

Different colour boxes indicate octree on different levels. After deletion of outliers in each node, the boxes are removed and the removed outliers, are analysed in the report.

Point cloud feature detect algorithm.

The algorithm uses PCA. After removing points, point cloud is neater and we can now detect features like edges. Due to large point cloud, data must be sampled for fast processing. PCA cannot be implemented on the whole point cloud as it would take a long time to process, and clients want faster procedures. After sampling is carried out, PCA is applied on this sampled data where normal is then extracted. A threshold is then implemented to remove outliers. The planes extracted, go through iterations till the number of points on the plane are stable. This results in ‘Best fit plane’.

3.4 Data Modelling: PCA

Data modelling is the design of the structure of data and data relationships. It is a process that involves some level of analysis. Data modelling makes adaptability to a new territory, easier and makes it simplifies decision making, in complicated situations. A data model should be logical and easy to comprehend. This technique is one useful way to better understand the connections in a dataset.

Data modelling can further be defined as an official depiction of the content and structure of data. It also helps to determine how data is stored and accessed thereby assisting users to map their data landscape.

Data modelling helps to prevent redundancies and gaps and it also gives a feeling of trust that the dataset is trusted and reliable.

Here we will look at Principal component analysis (PCA) for modelling data for landscape design.

PCA is a popular dimensionality reduction technique for large datasets and one of the most accurate statistical methods out of the existing methods. It involves extracting important components in a large dataset that contains numerous multiple components, with minimal information loss. Although the level of accuracy of a dataset is reduced, it produces a simpler dataset, which is easier to analyse and visualise. It can be used with 3D or higher dimensional data.

PCA is used for creating predictive models. Models are and approximates the system from where the data emerged. When applied to data, it produces three components.

- The first component represents the largest in any data set.
- The second component and
- The third component which has the least variance in any data set.

Before PCA modelling is carried out, the data is pre-processed and fed to the algorithm, to build the latent variable model. Worthy of note is that better models always emerge from transformed variables.

The steps involved in PCA are:

- Data standardization
- Calculate covariance.
- Obtain eigenvectors and eigenvalues from the covariance or correlation matrix as these are the most important part of PCA.
- Sort eigen values in descending order.
- Project data points to those eigen values.

4 Critical Analysis and Evaluation of the Applications of the Technique

4.1 Applications of Requirement technique

Data requirements analysis can be used in:

1. Articulating a clear understanding of data needs of all consuming business processes,
2. Identifying relevant data quality dimensions associated with those data needs,
3. Assessing the quality and suitability of candidate data sources,
4. Aligning and standardizing the exchange of data across systems.
5. Implementing production procedures for monitoring the conformance to expectations and correcting data as early as possible in the production flow, and
6. Continually reviewing to identify improvement opportunities in relation to downstream data needs.

4.2 Applications of 3D laser scanners

Laser scanning have a variety of uses which range from personal to Government sponsored projects like road surveys amongst others.

- Laser scanners are time efficient and minimise errors, thereby making it a preferred method for mobile mapping, surveys, scanning of structures and their interiors, by researchers.
- It is useful in surveying, reverse engineering, mining, and majority of civil engineering related projects.
- Laser scanners are useful in road surveys, traffic construction analysis as they reduce turnaround time remarkably.
- For infrastructure, 3D laser scanning, is used for the creation of digital terrain models and topographical measurements.

4.3 Applications of Data cleaning

Data cleaning is a necessary to avoid getting inaccurate results. Without proper cleaning of a dataset. Data analysis processes can be skewed thereby resulting in a result bias.

Data cleaning helps ensure that information always matches the correct fields while making it easier for business intelligence tools to interact with data sets to find information more efficiently. One of the most common data cleaning examples is its application in data warehouses.

Outlier analysis has numerous applications in a wide variety of domains such as the financial industry, quality control, fault diagnosis, intrusion detection, web analytics, and medical diagnosis.

4.4 Applications of PCA

PCA is used in producing highly efficient machine learning models by making use of dimensionality reduction to reduce system complexity.

- Facial recognition

- Image compression
- Used in finding patterns in data finance, data mining, psychology and bioinformatics.
- It is used in data visualisation.
- Data reconstruction
- Dimension reduction
- Recovering of data manifolds

5 Critical Review of the Commercial and Open-source Software

5.1 Software for BPM

Commercial Software:

Using Pipefy's business process management (BPM) software proffers an incredibly fast way to manage and enhance daily run procedures. **It creates an open, end-to-end operation to improve team efficiency, from input to delivery. It uses** enterprise-class process automation thereby enabling businesses focus on strategic initiatives and activities.

With this software, Business processes such as employee onboarding, job opening, purchasing are carried out in a less complex manner thereby increasing the output and efficiency of the business team.

Without having to deploy IT, a lot of things can be achieved such as speed and visibility increase, thereby producing higher quality outcomes.

It is fast growing brand driven by organizations needs for improved solutions in the way businesses are run. Pipefy software is utilised in various countries and by top leading firms like Visa, IBM, Volvo, Santander, Kraft Heinz and others.

Advantages of Pipefy

The software can be moulded according to the users' needs and it simplifies daily tasks.

Open-Source software: Bizagi Modeler

Free, intuitive, and collaborative business process mapping software used by millions of people across the globe. Bizagi Modeler enables organizations to create and document business processes in a central cloud repository to gain a better understanding of each step and identify process improvement opportunities to increase organizational efficiency.

To start using it, one needs to create or sign into the Bigazi account, get Bigazi modeler and proceed with mapping.

Advantages of Bigazi modeler

1 Unlimited individual use

2 Manage and organize process diagrams through folders.

3 Create and publish process documentation (PDF, Microsoft Word, Microsoft Excel, Wiki)

4 Import existing Microsoft Visio, IBM Bluewoks, XPDL, and BPMN diagrams

Cloud process repository (10MB)

5.2 Software for 3D scanner

Commercial Software: Agisoft MetaShape

Agisoft MetaShape is an independent 3D laser scanning software which transforms digital images by photogrammetry. It also produces 3D geographic data which are utilised in GIS applications, production of visuals, object measurement and other aspects. It is a commercial software which is run on Windows, Linux and Mac OS X. possessing a has a wide range of capabilities.

This product is an innovative software solution that has a (is) structured to produce tailored results, which are dependent on machine learning techniques for editing and analytic procedures.

Images are processed from RGB or including multi-camera systems, into the high-value geographic data in the form of compact point clouds, textured polygonal models, georeferenced true orthomosaics and DSMs/DTMs. Further post-processing removes noise for further analysis including classification of dense point clouds.

Advantages of Agisoft MetaShape

1. It has an advanced level of accuracy as it gives precise results
2. Fully automated and intuitive workflow
3. It has GPU acceleration for speedy processing and Network processing for large objects.
5.It has Agisoft Cloud for processing, visualization & sharing of the results
6. Reasonably powerful Standard edition for art projects
7. Stereoscopic measurements for precise feature extraction

Open-Source Software: MakerScanner

The MakerScanner is an opensource 3D-scanner which is a great accompaniment to the MakerBot or other 3D printers. Every one of the scanner's plastic parts can be printed on a MakerBot or other 3D printer.

It operates by energy projection into a scene to enable it carry out ranging. Its primary mode of operation can be likened to that of stereo vision. A laser pointer (scanning to make a vertical line) and a camera are offset a base distance B. The shorter the proximity of the objects to the sensor, The closer the appearance of the laser line to the camera image’s edge. The range to each point, can be determined, with a combination of maths and accurate standardization.

Advantages

1. It is relatively affordable

2. It has fast

3. It has a high level of accuracy

4. Beyond image digitization, minimal processing is required in contrast to the conventional stereo sensor which have the need for pixel correlation.

5.3 Software for removing outliers.

Commercial Software: 3DReshaper

3DReshaper is a commercial software, which can be used in various ways and has easy manageability. It processes point clouds from different sources ranging from 3D scanners to UAVs, and for various data fields, such as Architecture, Ship building, Geology, Quarry, Digital Terrain Modelling, Mine, and Civil Engineering.

Point cloud preparation is often the most important stage to handle in order to save time with the subsequent steps (i.e., meshing). (3D Reshaper, 2020)

That is why 3DReshaper provides a complete range of simple but powerful functions to process point clouds like: There is an exhaustive range of elementary but potent functions contained in this software for point cloud processing. They include:

…Import without real limit of the imported number of points
Clever reduction to keep best points and remove points only where density is the highest
Automatic segmentation (according to the scanning positions, distance, real or inspected colors)
Automatic (angle, inspection values) or manual separation and cleaning
Extraction of best points evenly spaced, density homogenization
Automatic noise measurement reduction
Colors according to a given direction
Fusion
Registration, Alignment and Best Fit

- 3D comparison with a mesh or a CAD model
- Planar sections
- Best geometrical shapes extraction (planes, cylinders, circles, spheres, etc.)
- Several representation modes: textured, shaded, intensity (information depending on the imported data)
- etc... (3D Reshaper, 2020)

Open-Source Software: CloudCompare

CloudCompare is an open-source software which uses an octree format. It is not just Windows, but Linux and macOS based also.

…CloudCompare is a 3D point cloud (and triangular mesh) processing software. It has been originally designed to perform comparison between two dense 3D points clouds (such as the ones acquired with a laser scanner) or between a point cloud and a triangular mesh. It relies on a specific octree structure dedicated to this task. Afterwards, it has been extended to a more generic point cloud processing software, including many advanced algorithms (registration, resampling, color/normal/scalar fields handling, statistics computation, sensor management, interactive or automatic segmentation, display enhancement, etc.). (Anon., n.d.)

5.4 Software for PCA

Commercial Software: MiniTab

MiniTab is a commercial software brand that can be used for data analyses. It offers an easy and useful way to store the statistical information, use that data, identify trends and patterns, and then deduce answers to the current issues. It is quite popular and used by many businesses as it provides a quick and effective solution when it comes to analysing data. It is a top-rated statistical software used for quality improvement and statistics education all over the globe.

Advantages

1. The software is easy to use and has assistant feature.

2. It can be used to teach statistics in higher institutions

Open-Source software: Past software

Past is an opensource software for implementing scientific data analysis. It is also inclusive of features for time series, plotting, univariate and multivariate statistics, ecological and spatial analysis, etc.
According to (LO4D.com, 2012) some of its features and highlights include:

1. A spreadsheet-type data entry form

2. Both interactive user interface and scripting

3. Graph, scatter, 3D scatter, bubble, histogram, kernel density estimation, box, percentile, ternary, survivorship, spindle, matrix, surface, and normal probability plots

4. Curve fitting: Linear (ordinary least squares, Reduced Major Axis, Major Axis, robust) with bootstrapping and permutation, Durbin-Watson and Breusch-Pagan tests, Generalized Linear Model including logit regression, lin-log (exponential), log-log (allometric), polynomial, logistic, von Bertalanffy, Michaelis-Menten, sum-of-sines, smoothing splines, LOESS smoothing, Gaussian (species packing), multiple multivariate regression, logarithmic spirals

6 Conclusion

The case study used for this report is the McCarthy Taylor Systems Limited and of the existing Data Analysis methods, the highlighted ones, for the purpose of this report, are requirement, collection, cleaning, and modelling techniques. The applications of these techniques are stated, alongside their corresponding software's. Both the proprietary and free tools are highly effective for these techniques but, this does not mean that other software's cannot be used as there are numerous out there.
Data Scientists are welcome to explore other existing data analyses methods and software packages for business purpose.

7 References

3D Reshaper, 2020. *3D Reshaper The 3D scanner software.* [Online]
Available at: https://www.3dreshaper.com/en/software-en/3dreshaper-en/point-cloud-processing
[Accessed 17 May 2021].

Agisoft, 2021. *Discover intelligent photogrammetry with Metashape.* [Online]
Available at: https://www.agisoft.com/
[Accessed 18 May 2021].

Analytics Vidhya, 2016. *PCA: A Practical Guide to Principal Component Analysis in R & Python.* [Online]
Available at: https://www.analyticsvidhya.com/blog/2016/03/pca-practical-guide-principal-component-analysis-python/
[Accessed 3 May 2021].

Anon., n.d. *CloudCompare.* [Online]
Available at: https://www.danielgm.net/cc/
[Accessed 17 May 2021].

Bizagi , 2021. *https://www.bizagi.com/en/platform/try-modeler.* [Online]
Available at: https://www.bizagi.com/en/platform/try-modeler
[Accessed 19 May 2021].

CompareCamp, 2020. [Online]
Available at: https://comparecamp.com/minitab-review-pricing-pros-cons-features/#:~:text=The%20main%20benefits%20of%20Minitab,without%20advanced%20knowledge%20of%20statistics.
[Accessed 16 May 2021].

dgirardeau, e. a., 2021. *CloudCompare.* [Online]
Available at: https://github.com/cloudcompare/cloudcompare
[Accessed 10 May 2021].

formplus, 2021. *Data Cleaning: 7 Techniques + Steps to Cleanse Data.* [Online]
Available at: https://www.formpl.us/blog/data-cleaning
[Accessed 15 May 2021].

G2, 2021. *Pipefy.* [Online]
Available at: https://www.g2.com/products/pipefy/pricing
[Accessed 18 May 2021].

Geeksforgeeks, 2018. *Principal Component Analysis with Python.* [Online]
Available at: https://www.geeksforgeeks.org/principal-component-analysis-with-python/
[Accessed 3 May 2021].

GetApp, 2010-2021. *Pipefy Reviews.* [Online]
Available at: https://www.getapp.com/operations-management-software/a/pipefy/reviews/
[Accessed 18 May 2021].

GoodFirms, 2021. *PAST.* [Online]
Available at: https://www.goodfirms.co/software/past/core-features
[Accessed 18 May 2021].

GreyCampus, n.d. *Minitab.* [Online]
Available at: https://www.greycampus.com/opencampus/minitab/introduction-on-minitab#:~:text=Minitab%20is%20a%20software%20product,answers%20to%20the%20current%20issues.
[Accessed 18 May 2021].

IBM Cloud Education, 2020. *Data Modeling.* [Online]
Available at: https://www.ibm.com/cloud/learn/data-modeling
[Accessed 15 May 2021].

Ko, R. K. L., 2009. A Computer Scientist's Introductory Guide to Business Process Management (BPM). *The ACM Magazine for Students,* 15(4), p. 11.

LeanIX, 2020. *THE DEFINITIVE GUIDE TO DATA MODELLING.* [Online]
Available at: https://www.leanix.net/en/data-models
[Accessed 14 May 2021].

LO4D.com, 2012. *https://past.en.lo4d.com/windows#screenshot_header.* [Online]
Available at: LO4D.com
[Accessed 20 may 2021].

MAKER SCANNER , n.d. *OPEN SOURCE 3D SCANNING.* [Online]
Available at: http://www.makerscanner.com/index.htm
[Accessed 18 May 2021].

Marie-Julie Rakotosaona, e. a., 2019. POINTCLEANNET: Learning to Denoise and Remove Outliers from Dense Point Clouds. 1 January.

Metashape, 2021. *Intelligent photogrammetry.* [Online]
Available at: https://www.agisoft.com/pdf/metashape_presentation.pdf
[Accessed 18 May 2021].

Nazar, B.-e., 2017. *Analytical Tools in Landscape Architecture.* [Online]
Available at:
https://www.researchgate.net/publication/320058817_Analytical_Tools_in_Landscape_Architecture
[Accessed 16th May 2021].

pipefy, 2021. *BUSINESS PROCESS MANAGEMENT.* [Online]
Available at: https://www.pipefy.com/process-management-software/
[Accessed 20 May 2021].

Pratt, M. K., 2020. *business process management (BPM).* [Online]
Available at: https://searchcio.techtarget.com/definition/business-process-management
[Accessed 20 May 2021].

ProjectPro, 2021. *Principal Component Analysis Tutorial.* [Online]
Available at: https://www.dezyre.com/data-science-in-python-tutorial/principal-component-analysis-tutorial#:~:text=Applications%20of%20Principal%20Component%20Analysis,%2C%20bioinformatics%2C%20psychology%2C%20etc.
[Accessed May12 2021].

Qlik, 1993-2020. *What is Data Modeling.* [Online]
Available at: https://www.qlik.com/us/data-modeling
[Accessed May 2021].

simplilearn, 2020. *What Is Requirement Analysis: Overview, Applications, Techniques and Top Tools Used.* [Online]
Available at: https://www.simplilearn.com/what-is-requirement-analysis-article
[Accessed 5 May 2021].

sisense, 2021. *Data Cleaning.* [Online]
Available at: https://www.sisense.com/glossary/data-cleaning/#:~:text=Data%20cleaning%20helps%20ensure%20that,its%20application%20in%20data%20warehouses.
[Accessed 21 May 2021].

tableau, 2003-2021. *Data cleaning: The benefits and steps to creating and using clean data.* [Online]
Available at: https://www.tableau.com/learn/articles/what-is-data-cleaning
[Accessed 3 May 2021].

Tripathi, A., 2019. *A COMPLETE GUIDE TO PRINCIPAL COMPONENT ANALYSIS – PCA IN MACHINE LEARNING.* [Online]
Available at: https://ashutoshtripathi.com/2019/07/11/a-complete-guide-to-principal-component-analysis-pca-in-machine-learning/
[Accessed 12 May 2021].

Visual Paradigm, 2020. *Requirement Analysis Techniques.* [Online]
Available at: https://www.visual-paradigm.com/guide/requirements-gathering/requirement-analysis-techniques/
[Accessed 14 May 2021].