Aleksandr Stepanov

data analytics  
 solution

REPORT

Table of Contents

[Introduction 2](#_Toc85364436)

[Business problem 2](#_Toc85364437)

[Importance of a Big Data 3](#_Toc85364438)

[Benefits & Advantages of Big Data Analytics 3](#_Toc85364439)

[Data analysis 4](#_Toc85364440)

[Collection 5](#_Toc85364441)

[Data quality 5](#_Toc85364442)

[Security and Privacy 6](#_Toc85364443)

[Data cleaning 6](#_Toc85364444)

[Data Mining 7](#_Toc85364445)

[Data visualization 7](#_Toc85364446)

[Summary 8](#_Toc85364447)

[Appendix 9](#_Toc85364448)

[References 11](#_Toc85364449)

## Introduction

Kathmandu is a transnational chain of retail stores. Was found in New Zealand in 1987 and it is a part of Holdings Limited. Kathmandu specialise in sustainably made Travel Gear, Adventure Gear, Outdoor Gear, Exercise Gear. (About Us, 2021).

In (Annual Report 2020, p. 10) Kathmandu reported the following information about performance of Holdings for 2020 in NZD:

* $426.4 m - total sales
* $80 m -online sales
* 67% growth year on year
* 18.5% of direct-to-consumer sales

There are 47 stores and 1 distribution centre in New Zealand. Australia has 116 stores 1 distribution centre (Sustainability Report , 2018, p. 10). Another centre is under consideration in Auckland, New Zealand. Data analysis can help make the right decision whether to open a new Distribution Centre.

# Business problem

The purpose of this report is to help Kathmandu find Big Data-based solution for their business problem.

There are 13 shops and no Distribution centre in Auckland (FIND A STORE KATHMANDU, 2021) with a population of 1,6 million of people (Auckland Popula, 2021), and 12 for the whole South Island. There are five shops in the biggest South Island city - Christchurch and one main distribution centre in Woolston.

# Importance of a Big Data

Big Data helps to make informed business decisions, improves understanding of customer behaviour and their purchasing pattern. It helps a company by cost saving, time saving, market understanding, offer marketing insights (Techvidvan, 2021). To get all the benefits of Big Data company must apply Big Data Analysis.

# Benefits & Advantages of Big Data Analytics

Applying this method will help to understand customer behaviour in a store or website.

Companies that implement Big Data analytics have access to benefits such as:

* Improve Customer Experience
* Quicker and Better Decision Making Within Organizations
* Product Development and Innovations
* Risk Management (What is big data analytics, 2021)

Based on them, company will know precisely what people will buy, it is possible to increase sales:

* by offering relevant related products while shopping.
* arranging sales and discounts on goods that are relevant at this time for most customers.
* letting them know about personalized discounts and offers.

## Data analysis

The increasing size and complexity of data has led to a need for inspection, data cleaning, transformation, and data modelling to gain insights from the data in order to make better decisions. This process is known as data analysis (Data-flair, 2021).

Figure 1 Data Analysis Process (Cho, 2015)

Diagram

Description automatically generated

1. Acquiring – gathering a data in one place.
2. Formatting - formatting the incoming data into an analysable form.
3. Ranging – data exploration, filtering, structuring, and querying.
4. Mining – pattern discovery and extraction of a useful information
5. Visualizing – representing of the meaningful information

During Interacting process Company can apply changes during this step and the data can be reprocessed to mining or ranging. Through this circle process the visualisation of a Big Data become more accurate, which leads to finding the right solution.

# Collection

To start the Big Data, process the data must be collected in a correct way. All raw pieces of information (datum) must be capable of being stored and moved. Tables of raw data should have a meaning variables for a company (Computerworld, 2021).

After raw data has been collected, it is sent to **Data Warehouse** - a system that contains the entire collected data for further simplified use. The data that is stored in Data Warehouse can be extracted and analyzed, name of this data - Historical Data. It does not store current information, nor is it updated in real-time (Panoply, 2021).

Kathmandu LTD can store information about purchases in store and online. For example: name of a product, quantity, date of a purchase(season). Moreover, company stores personal information of a members of a Summit Club, such as Name, Surname, date of birth and Email of a member (Join Summit Club, 2021)

# Data quality

To find out about reliability of a data quality control should be carried out.

For the best and most accurate result it is highly recommended that company apply quality assurance practices such as establishing standards and criteria for their data. Failure to follow these practices can lead to inaccurate, incomplete, or redundant data, what can be reason for making a wrong data-based decision and lead to loss of many (Marketing evolution, 2021).

To avoid bad data according to (Suer, 2021) **the collected data for the analysis must be**

* Consistent – data has no conflicts in information
* Accurate - ensuring values are correct and close to the reality of the results
* Formatted - data entry formats must be consistent
* Time framed - **data sets are up to date**
* Valid data is authentic and correct

# Security and Privacy

Security for Big Data projects is not just about making information accessible. The data that serve as a source for analysis contains sensitive information - personal data. Violation of the confidentiality of working with such data can turn into serious problems, including fines from regulators, customer loss, loss of market capitalization.

Another significant challenge of Big Data projects is ensuring the integrity of both the analyzed data and the results obtained during their processing, which are of commercial value (Gemalto, 2021).

There are few security tools that can help to prevent it, it is important that company apply them:

* **Encryption**
* **User Access Control**
* **Intrusion Detection and Prevention**
* **Centralized Key Management**

Customers do not always know what is happening with their data and where the data is stored**. Company needs to create a process approach to analysing and processing data, as well as automating processes related to ensuring the security of Big Data as part of common practice** (Maayan, 2021).

# Data cleaning

Missing numbers, misplaced entries, and typographical errors are all common data flaws. In some cases, data cleansing necessitates the filling in or correction of particular values, while in others, the entries must be eliminated entirely (talend, 2021).A clean dataset is vital for accurate data analysis. Data cleansing, or scrubbing data, is the first step in the overall data preparation process. It involves analyzing, identifying, and correcting messy data. Cleansing data involves filling in missing values, identifying errors, and ensuring that all information is arranged properly in rows and columns. The first step to making strategic business decisions is to clean organizational data thoroughly. Data cleaning paves the way for accurate, efficient, and effective data analysis (Trifacta, 2021)

# Data Mining

Data mining is the technique of locating anomalies, patterns, and correlations inside data to predict outcomes (Sas, 2021).

* **Cluster** Grouping similar records together.
* **Regression** - It forecasts the likelihood of an occurrence with only two possible outcomes. predict a range of numeric values.
* **Association** - Detecting relationships between records.
* **Sequential** - repeating pattern on one or several successive transactions of many inputs, identify associations over time sequences
* **Anomaly Detection** - Outlier Detection,exception
* **Prediction** – forecasting

Adherence to these techniques will help the company to identify significant trends, patterns in the behavior of customers in relation to purchases, will help to find new customers, which will help to make smart market decisions. Proper usage of data mining process will give a company the power of knowledges, what will lead to success and right decision about opening a new distribution center in Auckland (Korobka, 2021).

# Data visualization

Data visualization is the process of translating large data sets and metrics into charts, graphs, and other visuals. A data visualization will help company see and understand trends, outliers, and patterns in their data (Tableau, 2021). The most **common** types of data visualization are:

* Charts
* Tables
* Graphs
* Maps
* Infographics

Figure 2 Examples of visualizations

Diagram

Description automatically generated with medium confidence

Visual information is better perceived and allows quickly and effectively convey thoughts and ideas to the viewer. Physiologically, the perception of visual information is very important for humans. There are numerous studies confirming that:

* 90% of information a person perceives through sight
* 70% of sensory receptors are in the eyes
* about half of the neurons in the human brain are involved in the processing of visual information
* 19% less when working with visual data, the cognitive function of the brain is used, which is responsible for processing and analyzing information
* 17% higher productivity of a person working with visual information
* 4.5% better recall of detailed visual information (Why and how to use data visualization, 2021)

Kathmandu can use graphs to map visualizations of online purchases on the North Island and the South Island to reveal patterns that could lead to a decision to open a new distribution centre in Auckland. Chart may show the cost of delivery for the client and the company from the South centre to the North, there may also be a number of days in the chart showing waiting time for a customer. Waiting time might be a very important aspect for a customer as big delays may cost a company of losing a customer.

## Summary

To achieve the most favourable result, the company should complete all the steps of the date analytics and keep updating, gathering, and circling through the data.

We may conclude that Big Data assists companies in making smart decisions and better understanding their customers' needs.

By evaluating real-time data, data analysis can help Kathmandu LTD in achieving quick growth. It enables company to outperform their competitors and succeed.

Big Data analytics assist company in identifying inefficiencies and opportunities within retail business. It has a significant impact on the organization's development.

The use of Big Data analytics can help Kathmandu LTD:

* Unlock the strategic value of their assets and fully exploit them
* Predict market trends
* Predict future needs
* Make company more innovative and competitive

# Appendix

Figure 3 Total sales for 2020

A picture containing timeline

Description automatically generated

Figure 4 summit club membership

Graphical user interface, application

Description automatically generated

# References

*About Us*. (2021). Retrieved from Kathamndu Holdings: https://www.kathmanduholdings.com/about-us/

*Auckland Popula*. (2021). Retrieved from World Population Review: https://worldpopulationreview.com/world-cities/auckland-population

B, A. (2021, 10 12). *The data flow breakdown*. Retrieved from datamakesense: https://datamakessense.com/data-flow-breakdown/

Cho, W. (2015). Big Data Analysis With Interactive Visualisation . *Conference paper*, 3. Retrieved from Research gate: https://www.researchgate.net/publication/283853327\_Big\_Data\_Analysis\_with\_Interactive\_Visualization\_using\_R\_packages/link/5648753b08ae451880ae6150/download

Computerworld. (2021, 10 10). *Collections of Data*. Retrieved from Computerworld: https://www.computerworld.com/article/2588233/collections-of-data.html

Data-flair. (2021, 10 13). *Data Analytics*. Retrieved from data flair: https://data-flair.training/blogs/data-analytics-tutorial/

Gemalto. (2021, 10 12). *Gemalto Russia*. Retrieved from habr: https://habr.com/ru/company/gemaltorussia/blog/320214/

habr. (2021, 10 17). *Why and how to use data visualization*. Retrieved from habr: https://habr.com/ru/company/developersoft/blog/240325/

IBM. (2021, 10 04). *BIg Data Analytics*. Retrieved from IMB: https://www.ibm.com/analytics/hadoop/big-data-analytics

kathamndu. (2021, 10 04). *FIND A STORE KATHMANDU*. Retrieved from kathamndu: https://www.kathmandu.co.nz/stores

Kathmandu. (2021, 10 10). *Join Summit Club*. Retrieved from Kathmandu: https://www.kathmandu.co.nz/account/create-account

Kathmandu Holdings Limited. (2018). *Sustainability Report .* Retrieved October 29, 2019, from https://www.kathmanduholdings.com/wp-content/uploads/2018/10/Kathmandu\_susreport\_18\_lowres\_spreads-5.pdf

Kathmandu Holdings Ltd. (2020). *Annual Report 2020.* Retrieved from Kathmanduholdings.com: https://www.kathmanduholdings.com/wp-content/uploads/2020/10/KHL-Annual-Report-2020\_online.pdf

Korobka, M. (2021, 10 13). *REASONS WHY COMPANIES ARE OBSESSED WITH DATA MINING*. Retrieved from light-it: https://light-it.net/blog/reasons-why-companies-are-obsessed-with-data-mining/

Maayan, G. D. (2021, 10 12). *Big Data Security: Challenges and Solutions*. Retrieved from dataversity: https://www.dataversity.net/big-data-security-challenges-and-solutions/

Marketing evolution. (2021, 10 12). *What is Data Quality?* Retrieved from marketingevolution: https://www.marketingevolution.com/marketing-essentials/data-quality

ngdata. (2021, 10 17). *What is big data analytics*. Retrieved from ngdata: https://www.ngdata.com/what-is-big-data-analytics/

Panoply. (2021, 10 10). *What is a Data Warehouse*. Retrieved from Panoply: https://panoply.io/data-warehouse-guide/the-difference-between-a-database-and-a-data-warehouse/

SAS. (2021, 10 04). *BIG DATA*. Retrieved from SAS: https://www.sas.com/en\_nz/insights/big-data/what-is-big-data.html

Sas. (2021, 10 13). *data mining*. Retrieved from sas: https://www.sas.com/en\_nz/insights/analytics/data-mining.html

Suer, M. (2021, 10 12). *What Is Data Quality and Why Is It Important?* Retrieved from alation: https://www.alation.com/blog/what-is-data-quality-why-is-it-important/

(2018). *Sustainability Report .* Kathmandu Holding Ltd.

Tableau. (2021, 10 17). *What Is Data Visualization?* Retrieved from tableau: https://www.tableau.com/learn/articles/data-visualization

talend. (2021, 10 17). *What is Data Cleansing?* Retrieved from talend: https://www.talend.com/resources/what-is-data-cleansing/

Techvidvan. (2021, 10 13). *Why big data*. Retrieved from Techvidvan: https://techvidvan.com/tutorials/why-big-data/

Trifacta. (2021, 10 17). *How to Clean Data: Cleaning, Cleansing, and Scrubbing*. Retrieved from trifacta: https://www.trifacta.com/data-cleansing/

*What is Data Analytics?* (2021, 10 04). Retrieved from investopedia: https://www.investopedia.com/terms/d/data-analytics.asp