Aleksandr Stepanov

data analytics  
 solution

REPORT

Table of Contents

[Introduction 2](#_Toc82674986)

[Conclusions and Recommendations 3](#_Toc82674987)

[Appendix 4](#_Toc82674988)

[References 5](#_Toc82674989)

# 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.

# Resolving a 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 Big Data

Big data helps to make business decisions, improves understanding of customer behaviour and their purchasing pattern.

* You should then give an overview of the BI proces. Do NOT use the pyramid from the powerpoint, that is a different medium and not your own work.  Make your own diagram based on researched information.
* Then go into detail about each stage of the process, again based on researched information, but also applied to the scenario - e.g. for the first stage explain how raw data is stored in operational databases, and therefore what kind of raw data will Kathmandu be generating.
* You ARE giving them a plan as to how they could use BI/Analytics to use their big data to help them make the decision. This is an IT course so you are researching how you would use an IT solution to help solve a business problem, not researching business information to solve the problem itself.

# BI process

Diagram

Description automatically generated

* Data analytics is the science of analyzing raw data to make conclusions about that information.
* The techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.
* Data analytics help a business optimize its performance. (What is Data Analytics?)
* <https://www.investopedia.com/terms/d/data-analytics.asp>

Data analytics is the pursuit of extracting meaning from raw data using specialized computer systems. These systems transform, organize, and model the data to draw conclusions and identify patterns.

Big data analytics is the use of advanced analytic techniques against very large, diverse big data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.

What is big data exactly? It can be defined as data sets whose size or type is beyond the ability of traditional [relational databases](https://www.ibm.com/analytics/relational-database) to capture, manage and process the data with low latency. Characteristics of big data include high volume, high velocity and high variety. Sources of data are becoming more complex than those for traditional data because they are being driven by [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence), mobile devices, social media and the Internet of Things (IoT). For example, the different types of data originate from sensors, devices, video/audio, networks, log files, transactional applications, web and social media — much of it generated in real time and at a very large scale.

With big data analytics, you can ultimately fuel better and faster decision-making, modelling and predicting of future outcomes and enhanced business intelligence. As you build your big data solution, consider open source software such as [Apache Hadoop](https://www.ibm.com/analytics/hadoop), [Apache Spark](https://www.ibm.com/cloud/learn/apache-spark) and the entire Hadoop ecosystem as cost-effective, flexible data processing and storage tools designed to handle the volume of data being generated today. (IBM)

# 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 develop a process approach to analysing and processing data, as well as automating processes related to ensuring the security of big data as part of established practices**  (Maayan, 2021).

# Data cleaning

# Data Mining

# Conclusions and Recommendations

Plan:

Finish introduction

Security of big data or just data

Collecting a big data

Raw data

Process of big data, collection etc

* Data Mining.
* Data Collection.
* Data Storing.
* Data Cleaning.
* Data Analysis.
* Data Consumption.

# Appendix

A picture containing timeline

Description automatically generated

Graphical user interface, application

Description automatically generated

## References

# 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

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

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

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

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

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