**Jagannath International Management School**

**Vasant Kunj, New Delhi-110070**

(Affiliated to Guru Gobind Singh Indraprastha University, New Delhi)

Recognized u/s 2(f) by UGC & Accredited with ‘A’ Grade by NAAC

Participant of United Nations Global Compact, New York

ISO 9001:2015 Quality Certified



**Synopsis Report**

**on**

“Données”

**For Partial fulfilment of Bachelor of Computer Application**

Submitted To: Submitted By:

**Prof. Anisha Tandon** Group No. - 14

Assistant Professor-IT Department Aman Malik-05021402019

JIMS, Vasant Kunj, New Delhi Ritik Malik-04321402019

Gaurav Nailwal-05621402019

BCA-5(E)

**Données**

**TABLE OF CONTENTS**

* Introduction …………………………………...... 4
* Problem Statement ……………………………... 5
* Objective ………………………………………. 6
* Technical Specifications ……………………...... 7
  + Project Category
  + Programming Languages
* Scope of the Project ……………………….….... 8

**INTRODUCTION**

Data science is an emerging field of science that has multiple aspects – for one, it studies and examines a huge amount of data; for another, its branches extend in almost every field.

The data we work on is not simple; it is complex data that is structured in many layers. Data science is founded on three main components, and they are statistics, mathematics, and programming language.

Artificial intelligence encapsulates the concepts of all three fields and acts as the machinery or brain of data science. Data science uses techniques, procedures, algorithms, rules, and tools from all these three components and works as a unified mechanism to solve the complex problems that arise in the world around us.

Data is valuable, and so is the science in decoding it. Zillions of bytes of data are being generated, and now its value has surpassed oil as well. The role of a data scientist is and will be of paramount importance for organizations across many verticals.

Datasets are not very friendly to use with. But first, what is data set? A dataset is a collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer. The data is normally obtained from historical observations.

It is very challenging to find the relevant or the required data and is often difficult to work on it. To work with dataset the user would first have to set up an environment in which the dataset would be opened. Setting up the environment for dataset is also not an easy task. There only are a handful websites which allow us to get dataset and even more website which let us see the dataset without downloading them.

**PROBLEM STATEMENT**

Since there are only a few websites which allow us to use dataset online, we decided to build a website which will allow the user to see the dataset online, try and manipulate the dataset and if the user want, they could also download the dataset. All this will be done at one website named “Données”.

Our aim is to eliminate the problem of using dataset without downloading them. It is difficult to download each and every dataset we want to see. It requires an environment to use the downloaded dataset. Storing datasets is also a huge problem. It requires a lot of space. Our website will let the user to first see the dataset online, the user could work on the dataset without downloading it and if the user wants, they could also download the dataset with just one click.

We will solve the problem of:

1. Storing the dataset
2. See the dataset without downloading it.
3. Working on dataset without downloading it.

**OBJECTIVE**

The objective of this website is to explore, sort and analyse metadata from various sources in order to take advantage of them and reach conclusions to optimize business processes or for decision support.

In this website, we have shown data science to analyse the datasets with the help of some python libraries like NumPy, pandas, matplotlib etc. which helps in training of datasets, helps the user to conclude for future decision making or for predictions easily.

The key objective of Data Science is to extract valuable information for use in strategic decision making, product development, trend analysis, and forecasting. Data Science has several benefits like Identification and refining of target audiences, better risk analysis and recruit better in lesser time.

Mostly, we have the download the datasets for recognizing data for which we have to set environment setup in our system accordingly generally which takes a lot time for the user. But on our website, the user can see a wide variety of datasets online without downloading it on their system and no need of setup, the user could work on the dataset without downloading it. There is a separate option also to download dataset easily.

You can benefit from our excellent track record in using data to make a difference in both large and smaller businesses, including:

* Using predictive analytics and [machine learning](https://objectiveit.com/data-analytics/machine-learning/) to significantly increase the sales funnel.
* Improve and enhance customer segmentation.
* Reduce customer churn.
* Understand good and bad suppliers and customers.
* Improve geographic knowledge and much more.

Surely, it will help the user as it is less time consuming and more effectively to use and sort.

**TECHNICAL SPECIFICATIONS**

**Project Category**

EDA (Exploratory Data Analysis) and Data Science

**Programming Languages:**

**Frontend-**

* **Hypertext Markup Language:** It is the standard markup language for documents designed to be displayed in a web browser.
* **Cascading Style Sheet:** It is a style sheet language used for describing the presentation of a document written in a markup language such as HTML and is a cornerstone technology of the World Wide Web.
* **JavaScript:** It is a high-level, just-in-time compiled, multi-paradigm programming language that conforms to the ECMAScript specification. It has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions.
* **Bootstrap:** Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development.

**Backend-**

* **Php:** Hypertext preprocessor is a widely-used open-source general-purpose scripting language that is especially suited for web development and can be embedded into HTML.

**Database-**

* **SQL:** Structured query language, computer language designed to eliciting information from databases.
* **Python:** Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation.

**\**

**SCOPE of the PROJECT**

* People can view uploaded Datasets.
* People can send request and the admins will try to upload those datasets.
* In the Discussion Forum, people can post any query.
* The admins can publish Dataset and their respected workbooks.
* Admin has full authority of the website to manage users, discussion forum etc.

**Final conclusions and future research-**

The findings and implications explained in this research enable new users to get an insight into data science and EDA. All the links provided under the datasets will help the user to clear his or her concepts and all the opensource forums will be linked to the dataset itself to keep the user updated about the project and find all the libraries that are useful for future reference.

According to IDC, by 2025, global data will grow to 175 zettabytes. Data Science enables companies to efficiently understand gigantic data from multiple sources and derive valuable insights to make smarter data-driven decisions. Data Science is widely used in various industry domains, including marketing, healthcare, finance, banking, policy work, and more. That explains why Data Science is important and hence we are working on this project so enthusiastically.