DAYANANDA SAGAR COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING

MINOR PROJECT- REPORT
AUG-2022 JAN-2023

Course Faculty: Prof. Nidhishree M S Course Name & code: 19CS7DLCBL – Big Data
Semester: 7th C Date: 22/11/2022

TITLE OF THE PROJECT	**BUSINESS SALES ANALYSIS**			
STUDENT NAME	SHAMSHER TIWARI	SHUBHAM SHRESHTH	VIPUL TIWARI	
USN	1DS19CS147	1DS19CS156	1DS19CS190	
INDIVIDUAL CONTRIBUTION				
GUIDE	PROF. NIDHISHREE M			
PROJECT ABSTRACT:	Over past ten years, industries and organizations didn't need to store and perform much operations and analytics on data of the customers. But around from 2005, the need to transform everything into data is much entertained to satisfy the requirements of the people. "Without big data analytics, companies are blind and deaf, wandering out onto the Web like deer on a freeway." Whether you are trying to improve customer loyalty and engagement, optimize your performance, or make pricing decisions, big data in marketing has proven to be an indispensable tool.			

DAYANANDA SAGAR COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING

	How Big Data is Transforming marketing and sales? – 1. Getting 360-degree view of their customers 2. Brand Awareness 3. Customer Agreement 4. Improved customer acquisition 5. Leverage real time data in cloud computing environments 6. Results in time and cost savings
PLATFORM USED (H/W & S/W TOOLS TO BE USED	Jupyter Notebook, Google Collab
INTRODUCTION	In this project we will use Python Pandas & Python Matplotlib to analyse about 12 months' worth of sales data. The data contains hundreds of thousands of electronics store purchases broken down by month, product type, cost, purchase address, etc. We start by cleaning our data. Tasks during this section include: • Drop NaN values from DataFrame • Removing rows based on a condition • Change the type of columns (to_numeric, to_datetime, astype) Once we have cleaned up our data a bit, we move the data exploration section. In this section we explore 5 high level business questions related to our data: • What was the best month for sales? How much was earned that month? • What city sold the most product? • What time should we display advertisements to maximize the likelihood of customer's buying product? • What products are most often sold together? • What product sold the most? Why do you think it sold the most?

DAYANANDA SAGAR COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING

	To answer these questions, we walk through many different pandas & matplotlib methods. They include: • Concatenating multiple csv's together to create a new DataFrame (pd.concat()) • Adding columns • Parsing cells as strings to make new columns (.str) • Using the .apply() method • Using groupby to perform aggregate analysis • Plotting bar charts and lines graphs to visualize our results • Labelling our graphs.
DESIGN	
PROJECT SOURCE CODE LINK (GITHUB/ GOOGLE DRIVE)	
CONCLUSION /FUTURE ENHANCEMENT	
UI SCREENSHOTS	