CONTENTS

> ? @ AB @ AC

**ABSTRACT** **........................................................................................... 2**

**CHAPTER 1........................................................................................... 3**

**INTRODUCTION**

* 1. DATA ANALYTICS…………..............................................................................3

1.11 TYPES OF DATA ANALYTICS………………………………………………..4

1.2 PROJECT AIMS AND OBJECTIVES ................................................................5

1.3 BACKGROUND OF PROJECT ........................................................................5

1.4 OPERATION ENVIRONMENT .........................................................................6

**CHAPTER 2 ....................................................................................... 8**

**SYSTEM DESIGN AND IMPLEMENTATION**

2.1 SOFTWARE AND HARWARE REQUIREMENTS .............................................8

2.2 LIBRARIES AND PACKAGES USED...................................................................9

2.3 DATASET.............................................................................................................9

2.4 SCREENSHOTS ............................................................................................... 10

**CHAPTER 3.......................................................................................... 14**

**CONCLUSION AND FUTURE SCOPE**

**CHAPTER 4 ........................................................................................ 15**

**REFERENCES**

**ABSTRACT**

This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2017 for a UK-based and registered non-store online retail. One of the most important information given by data envelopment analysis models is the cost, revenue and profit efficiency of decision making units (DMUs). Cost efficiency is defined as the ratio of minimum costs to current costs, while revenue efficiency is defined as the ratio of maximum revenue to current revenue of the DMU. It presents a framework where data envelopment analysis (DEA) is used to measure cost, revenue and profit efficiency with fuzzy data. In such cases, the classical models cannot be used, because input and output data appear in the form of ranges. When the data are fuzzy, the cost, revenue and profit efficiency measures calculated from the data should be uncertain as well. Fuzzy DEA models emerge as another class of DEA models to account for imprecise inputs and outputs for DMUs. Although several approaches for solving fuzzy DEA models have been developed, numerous deficiencies including the α-cut approaches and types of fuzzy numbers must still be improved. This scheme embraces evaluation method based on vector for proposed fuzzy model. This paper proposes generalized cost, revenue and profit efficiency models in fuzzy data envelopment analysis.

**CHAPTER 1**

**INTRODUCTION**

This chapter gives an overview about analysis of data sets using python, the aim, objectives, and background and operation environment of the system.

**1.1 DATA ANALYTICS.**

As the process of analyzing raw data to find trends and answer questions, the definition of data analytics captures its broad scope of the field. However, it includes many techniques with many different goals.The data analytics process has some key components that are needed for any initiative. By combining these components, a successful data analytics initiative will provide a clear picture of where you are, where you have been and where you should go.

Generally, this process begins with descriptive analytics. This is the process of describing historical trends in data. Descriptive analytics aims to answer the question “what happened?” This often involves measuring traditional indicators such as return on investment (ROI). The indicators used will be different for each industry. Descriptive analytics does not make predictions or directly inform decisions. It focuses on summarizing data in a meaningful and descriptive way.

The next essential part of data analytics is advanced analytics. This part of data science takes advantage of advanced tools to extract data, make predictions and discover trends. These tools include classical statistics as well as machine learning. Machine learning technologies such as neural networks, natural language processing, sentiment analysis and more enable advanced analytics. This information provides new insight from data. Advanced analytics addresses “what if?” questions.

The availability of machine learning techniques, massive data sets, and cheap computing power has enabled the use of these techniques in many industries. The collection of big data sets is instrumental in enabling these techniques. Big data analytics enables businesses to draw meaningful conclusions from complex and varied data sources, which has made possible by advances in parallel processing and cheap computational power.

**1.12 TYPES OF DATA ANALYTICS.**

Data analytics is a broad field. There are four primary types of data analytics: descriptive, diagnostic, predictive and prescriptive analytics. Each type has a different goal and a different place in the data analysis process. These are also the primary data analytics applications in business.

Descriptive analytics helps answer questions about what happened. These techniques summarize large datasets to describe outcomes to stakeholders. By developing key performance indicators (KPIs,) these strategies can help track successes or failures. Metrics such as return on investment (ROI) are used in many industries. Specialized metrics are developed to track performance in specific industries. This process requires the collection of relevant data, processing of the data, data analysis and data visualization. This process provides essential insight into past performance.

Diagnostic analytics helps answer questions about why things happened. These techniques supplement more basic descriptive analytics. They take the findings from descriptive analytics and dig deeper to find the cause. The performance indicators are further investigated to discover why they got better or worse. This generally occurs in three steps:

Identify anomalies in the data. These may be unexpected changes in a metric or a particular market.

Data that is related to these anomalies is collected.Statistical techniques are used to find relationships and trends that explain these anomalies.Predictive analytics helps answer questions about what will happen in the future. These techniques use historical data to identify trends and determine if they are likely to recur. Predictive analytical tools provide valuable insight into what may happen in the future and its techniques include a variety of statistical and machine learning techniques, such as: neural networks, decision trees, and regression.

Prescriptive analytics helps answer questions about what should be done. By using insights from predictive analytics, data-driven decisions can be made. This allows businesses to make informed decisions in the face of uncertainty. Prescriptive analytics techniques rely on machine learning strategies that can find patterns in large datasets. By analyzing past decisions and events, the likelihood of different outcomes can be estimated.

These types of data analytics provide the insight that businesses need to make effective and efficient decisions. Used in combination they provide a well-rounded understanding of a companies needs and opportunities.

**1.2 PROJECT AIMS AND OBJECTIVES**

The project aims and objectives that will be achieved after completion of this project are discussed in this subchapter. The aims and objectives are as follows:

* To increase efficiency and improve performance by discovering patterns in data.
* Working with data in various ways
* The primary steps in the data analytics process are data mining, data management, statistical analysis, and data presentation.
* The importance and balance of these steps depend on the data being used and the goal of the analysis.
* This involves extracting data from unstructured data sources. These may include written text, large complex databases, or raw sensor data.

**1.3 BACKGROUND OF PROJECT**

The applications of data analytics are broad. Analyzing big data can optimize efficiency in many different industries. Improving performance enables businesses to succeed in an increasingly competitive world. Data analytics is being used with great success in a number of different fields.One of the earliest adopters is the financial sector. Data analytics has a huge role in the banking and finance industries, used to predict market trends and assess risk. Credit scores are an example of data analytics that affects everyone. These scores use many data points to determine lending risk. Data analytics is also used to detect and prevent fraud to improve efficiency and reduce risk for financial institutions.The use of data analytics goes beyond maximizing profits and ROI, however. Data analytics can provide critical information for healthcare (health informatics), crime prevention, and environmental protection. These applications of data analytics use these techniques to improve our world.

Though statistics and data analysis have always been used in scientific research, advanced analytic techniques and big data allow for many new insights. These techniques can find trends in complex systems. Researchers are currently using machine learning to protect wildlife.The use of data analytics in healthcare is already widespread. Predicting patient outcomes, efficiently allocating funding and improving diagnostic techniques are just a few examples of how data analytics is revolutionizing healthcare. The pharmaceutical industry is also being revolutionized by machine learning. Drug discovery is a complex task with many variables. Machine learning can greatly improve drug discovery. Pharmaceutical companies also use data analytics to understand the market for drugs and predict their sales.The internet of things (IoT) is a field that is exploding alongside machine learning. These devices provide a great opportunity for data analytics.

IoT devices often contain many sensors that collect meaningful data points for their operation. Devices like the Nest thermostat track movement and temperature to regulate heating and cooling. Smart devices like this can use data to learn from and predict your behavior. This will provide advance home automation that can adapt to the way you live.The applications of data analytics are seemingly endless. More and more data is being collected every day — this presents new opportunities to apply data analytics to more parts of business, science and everyday life.

**1.4 OPERATION ENVIRONMENT**

**TABLE: 1.4**

|  |  |
| --- | --- |
| PROCESSOR | INTEL CORE PROCESSOR OR BETTER PERFORMANCE |
| OPERATING SYSTEM | WINDOWS VISTA ,WINDOWS7, UBUNTU |
| MEMORY | 1GB RAM OR MORE |
| HARD DISK SPACE | MINIMUM 3 GB |
| LANGUAGE USED | PYTHON |

### NOTEBOOK

The text editor will be used to type our program. Examples of few editors include Windows Notepad, OS Edit command, Brief, Epsilon, EMACS, and vim or vi.

Jupyter notebooks are interactive documents that contain code, narratives, plots. They are an excellent place for experimenting with code and data. Notebooks are easily shared, and the [2.6M notebooks](https://github.com/parente/nbestimate) on GitHub just tell how popular notebooks are.

Jupyter notebooks are great, but they often are huge files, with a very specific JSON file format. Let us introduce [Jupytext](https://github.com/mwouts/jupytext/blob/master/README.md" \t "_blank), a Jupyter plugin that reads and writes notebooks as plain text files: either Julia, Python, R scripts, Markdown, or R Markdown documents.

### Python Compiler

### The Python compiler package is a tool for analyzing Python source code and generating Python bytecode. The compiler contains libraries to generate an abstract syntax tree from Python source code and to generate Python [bytecode](https://docs.python.org/2/glossary.html#term-bytecode) from the tree.

### The [compiler](https://docs.python.org/2/library/compiler.html#module-compiler) package is a Python source to bytecode translator written in Python. It uses the built-in parser and standard [parser](https://docs.python.org/2/library/parser.html#module-parser) module to generate a concrete syntax tree. This tree is used to generate an abstract syntax tree (AST) and then Python bytecode.

### The full functionality of the package duplicates the built-in compiler provided with the Python interpreter. It is intended to match its behavior almost exactly. Why implement another compiler that does the same thing? The package is useful for a variety of purposes. It can be modified more easily than the built-in compiler. The AST it generates is useful for analyzing Python source code.

**CHAPTER 2**

**SYSTEM DESIGN & IMPLEMENTATION**

* 1. **SOFTWARE AND HARDWARE REQUIREMENTS**

This section describes the software and hardware requirements of the system

* + - 1. **SOFTWARE REQUIREMENTS**
* Operating system- Windows 7 or Windows 10 is used as the operating system as it is stable and supports more features and is more user friendly.
* Turbo C is used as text editor and C compiler as it easy to maintain and retrieve records by simple file operations which are in English language which are easy to understand and easy to write.
  + - 1. **HARDWARE REQUIREMENTS**

Intel core i5 5th generation is used as a processor because it is fast than other processors and provide reliable and stable and we can run our pc for longtime. By using this processor we can keep on developing our project without any worries.

Ram 2 GB is used as it will provide fast reading and writing capabilities and will in turn support in processing.

**2.2 PACKAGES AND LIBRARIES USED:**

**TABLE: 3.1.1**

|  |  |
| --- | --- |
| **Libraries & Pakage(s) used** | **Description** |
| numpy | Multi dimensional array support |
| pandas | Data Manipulation and Analysis |
| Matplotlib | Visualize information |
| OpenCV | Image Processing |
| Seaborn | Provides high level interface for drawing attractive and informative statistical graphics |

**2.3 DATA SET**

no\_data;Revenue;Cost;Profit

2010;100;-90;10

2011;120;-105.6;14.4

2012;130;-114.4;15.6

2013;233;-207.37;25.63

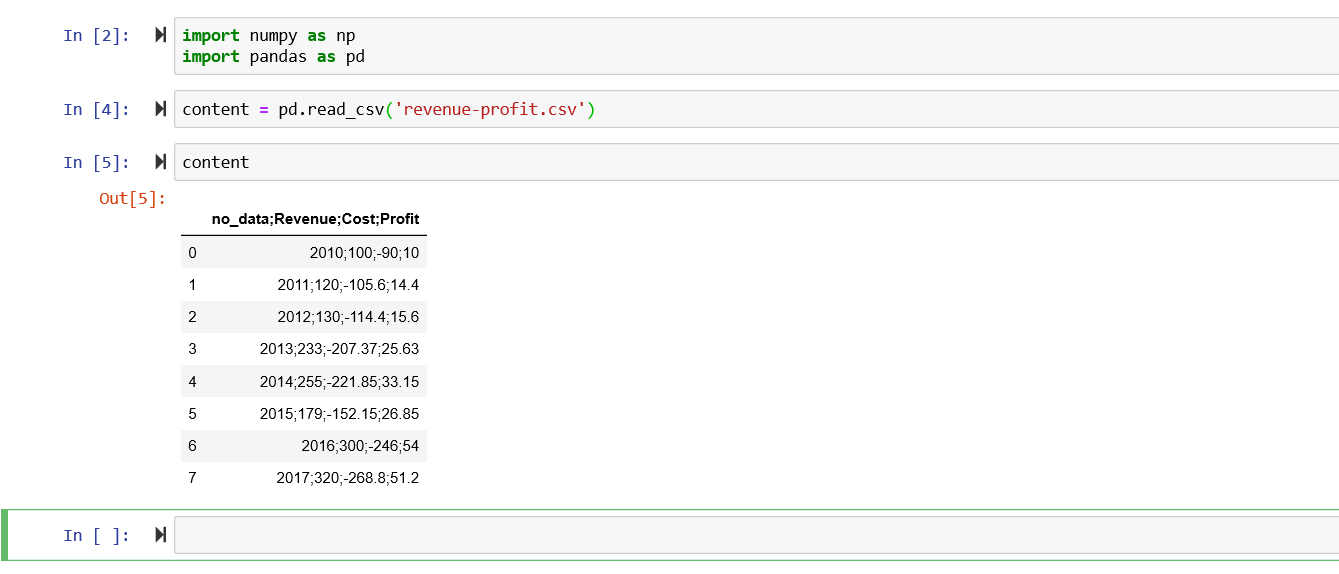
2014;255;-221.85;33.15

2015;179;-152.15;26.85

2016;300;-246;54

2017;

**2.4 SCREENSHOTS**

****

**Screenshot 2.4.1: Contents of the Data Set**

****

**Screenshot 2.4.2: Contents separation using ‘sep’ variable**



**Screenshot 2.4.3: Retrieving data using ‘content head()’**

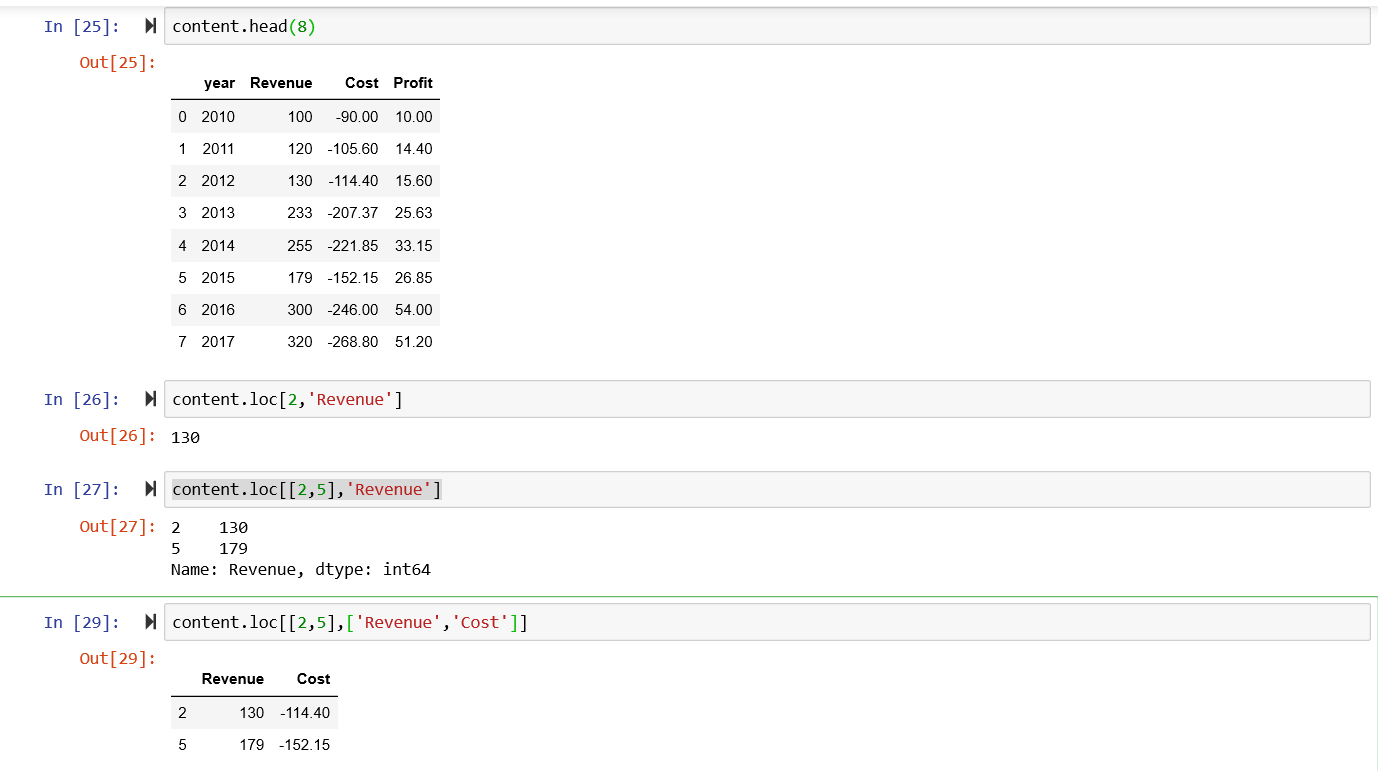


**Screenshot 2.4.4: Retrieving specified data using ‘single\_year’ and**

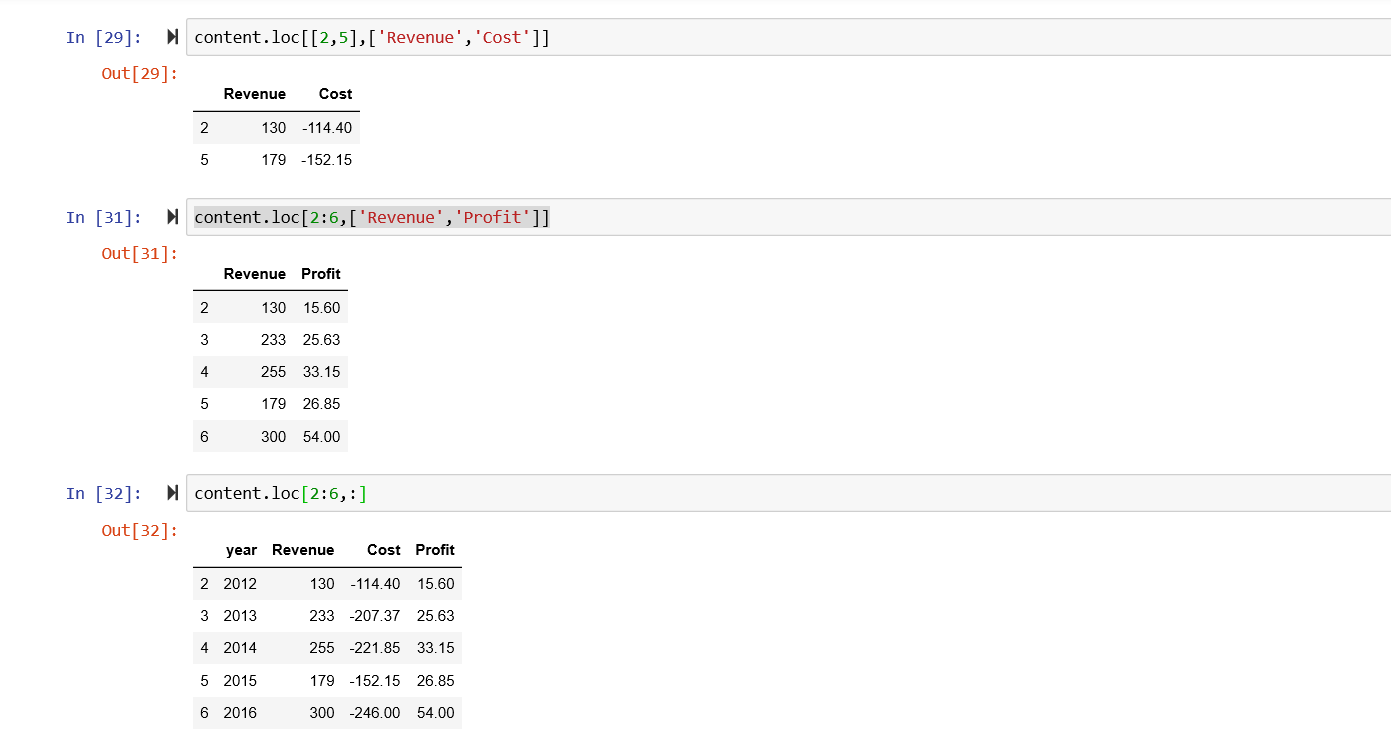
**‘single\_column’**



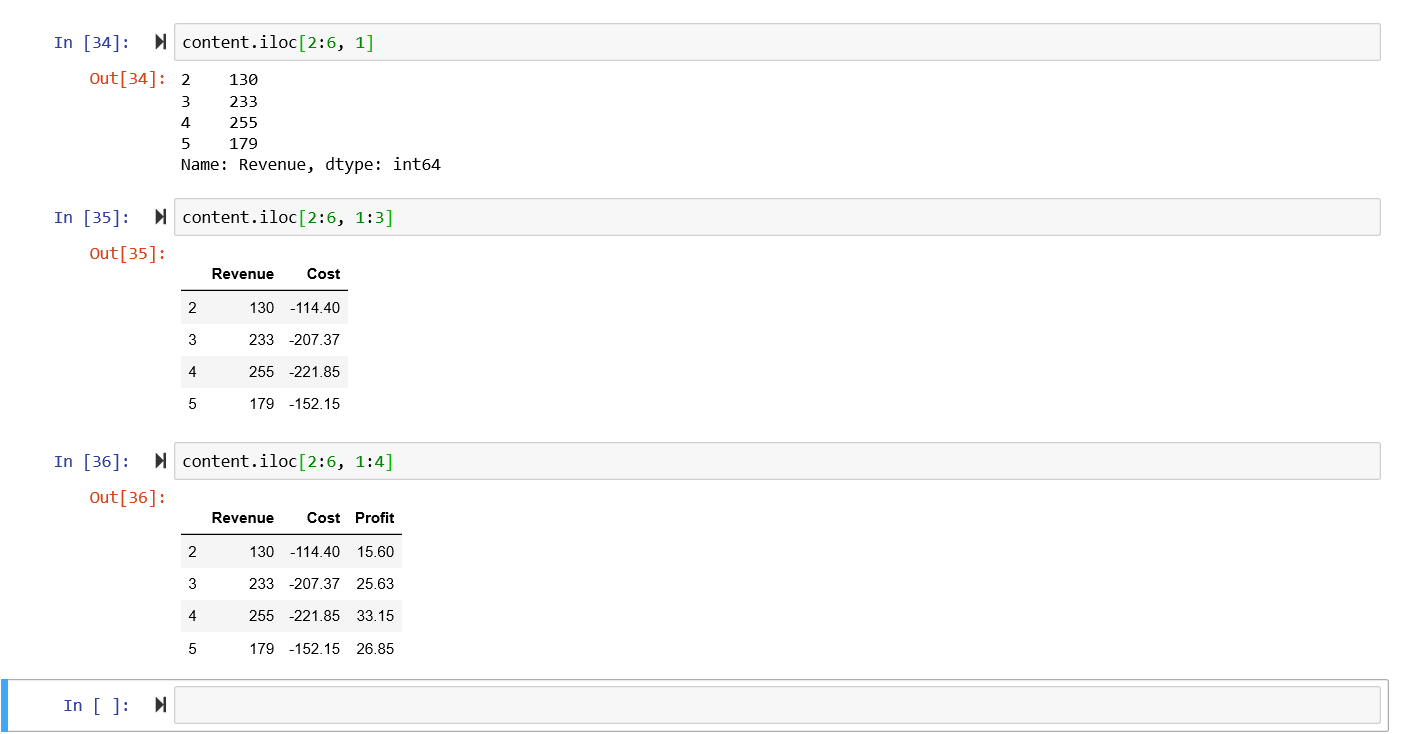
**Screenshot 2.4.5: Retrieving multiple column using ‘multiple\_column’**



**Screenshot 2.4.6: Retrieving specific location using ‘content,loc[]’**



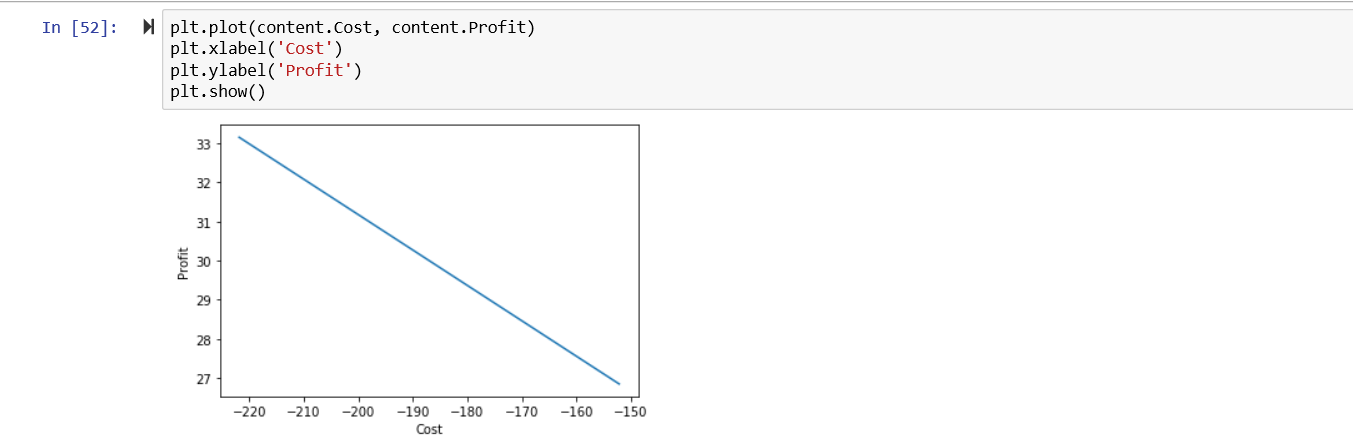
**Screenshot 2.4.7: Retrieving specific location using ‘content,loc[]’**



**Screenshot 2.4.8: Retrieving specific location using ‘content,iloc[]’**

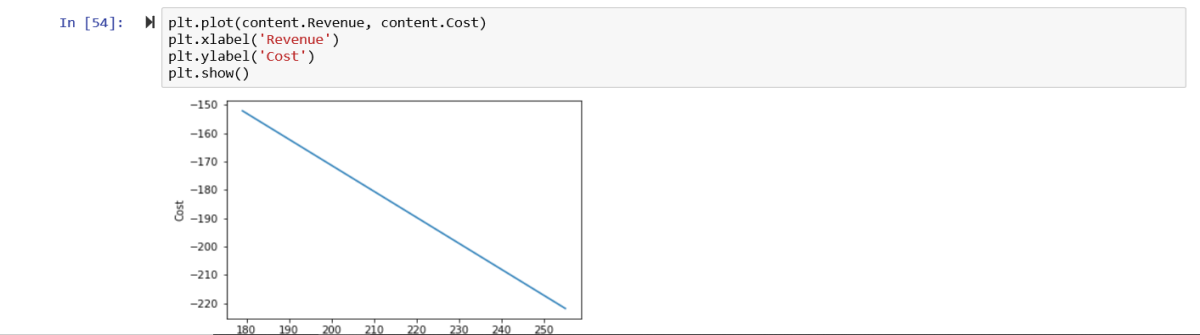


**Screenshot 2.4.9: Retrieving specific location using ‘content,iloc[]’**



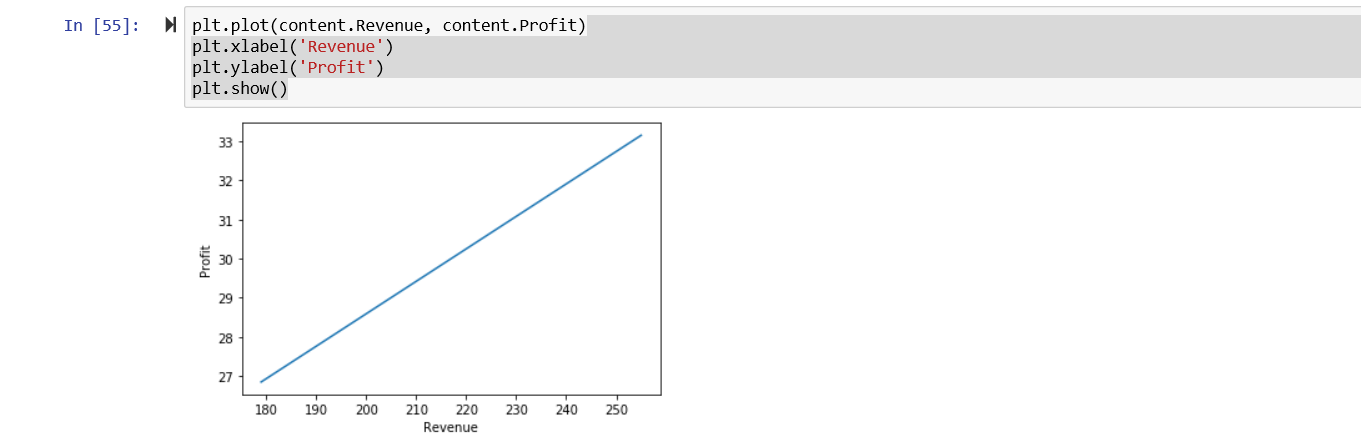
**Screenshot 2.4.10: Plotting a graph to retrieve ‘Cost’ and ‘Profit’**

**Statistics**



**Screenshot 2.4.11: Plotting a graph to retrieve ‘Revenue’ and ‘Cost’**

**statistics**

****

**Screenshot 2.4.12: Plotting a graph to retrieve ‘Revenue’ and ‘Profit’**

**statistics**

**CHAPTER 4**

**CONCLUSION AND FUTURE SCOPE**

Data analytics is a process through which data is cleaned, analysed and modelled using tools. This data is then used to derive insights. The insights are then used for business-related decision-making purposes. There are many techniques that data analysts use in different fields of work. In the world of business, Data analytics is used for making strategies to get the desired business results. Today, data analytics has become a big career option in India. As a result, big data analytics courses are in huge demand.Businesses have realised the importance of utilising big data analytics to maximize their profits. They know that it is vital for their growth and for the future health of their business. Today, major business decisions are taken by utilising the insights derived from data related to the organization or industry-related data. As competition increases and customers are flooded with choices, it has become important to move faster in the market and that too with accuracy.

Data analytics provides both speed and accuracy to business decisions. It provides accuracy as it is based on statistical models and hi-tech tools that help fine-tuning and analysing the data. This field also provides answers to present business problems as well as give a view of future trends. It is preparing the companies to make products for the future and aspire to connect with the customers of tomorrow.

 India is a popular destination for a lot of companies who outsource their work to other countries. This is due to the lower cost of operations and manpower in India. This is further aided by the skilled and English-speaking youth of India. Data analytics is one such field where outsourced opportunities are available in India. As a country teeming with young people and tremendous outsourced work coming in, the scope for this sector is big in India.

 Data analytics is the differentiator that provides companies with a competitive edge over others. It is a fast-growing branch of study which has a bright future in India. Organizations have realised their importance and investing in data analytics tools and technologies. Professionals and students are keen for a [career in data analytics](https://talentedge.com/spjimr-formerly-sp-jain-institute-of-management-and-research/big-data-analytics-certification/) owing to the career opportunities they might get. We can be sure that data analytics has a good future in India for years to come.

**CHAPTER 5**

**REFERENCES**

* <https://www.google.com/search?q=data+analytics&rlz=1C1CHBF_enIN865IN865&oq=data&aqs=chrome.0.69i59l2j69i57.3170j0j7&sourceid=chrome&ie=UTF-8>
* <https://www.google.com/search?q=data+set+revenue+profit+abstract&oq=dat&aqs=chrome.1.69i59l3j69i57j0j69i60l3.2779j0j7&sourceid=chrome&ie=UTF-8>
* <https://www.google.com/search?q=jupyter+text+editor&rlz=1C1CHBF_enIN865IN865&oq=ju&aqs=chrome.0.69i59j69i57j0l2j69i60l4.3403j0j9&sourceid=chrome&ie=UTF-8>