औद्योगिक प्रशिक्षण के लिए राष्ट्रीय संस्थान

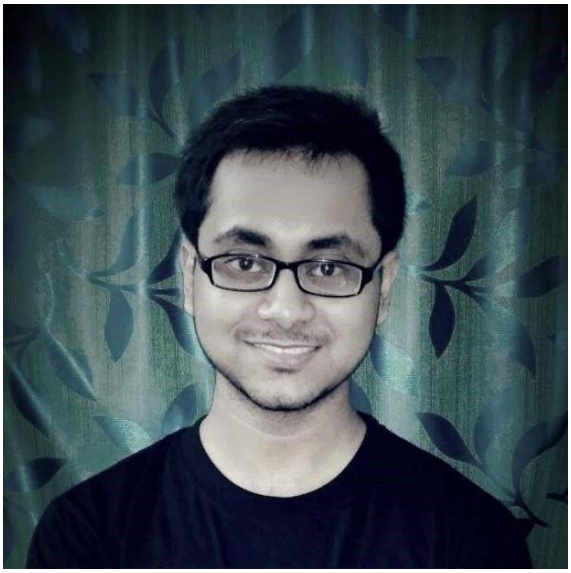
National Institute for Industrial Training

One Premier Organization with Non Profit Status | Registered Under Govt. of WB
Empanelled Under Planning Commission Govt. of India
Inspired By: National Task Force on IT & SD Government of India

National Institute for Industrial Training- One Premier Organization with Non Profit Status Registered Under Govt.of West Bengal,Empanelled Under Planning Commission Govt.of India , Empanelled Under Central Social Welfare Board Govt. of India , Registered with National Career Services , Registered with National Employment Services.

Innovative Project Topic → Analysis of Sales Data of an Ecommerce Company

Submitted to → Mr. Ankit Pramanik, Mr. Sayantan Chakraborty



Name → Sandipan Chowdhury

College → B.P. Poddar Institute of Management and Technology

Course → Computer Science and Engineering, B.Tech.

Session → 2016-2020

Email ID → sandipanchowdhury42@gmail.com

Phone number → 9123972867

Contents

Topic **Page Number**

Acknowledgement

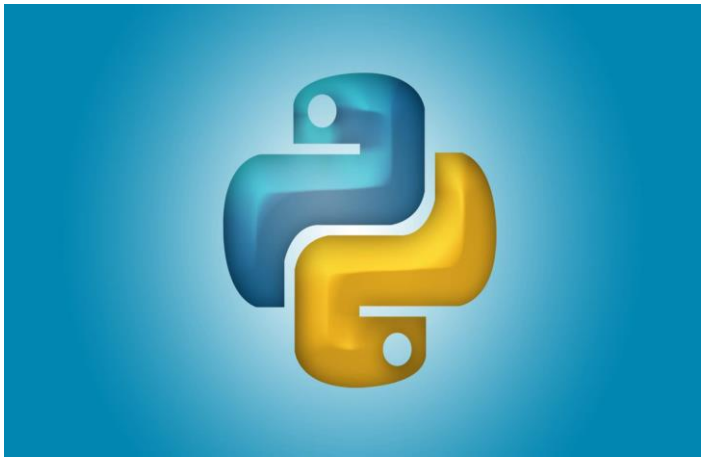
I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I am highly indebted to National Institute for Industrial Training for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. I would like to express my gratitude towards Mr. Ankit Pramanik, Mr. Sayantan Chakraborty for their support, co-operation and encouragement which helped me in completion of this project.

Python

Introduction → Python is an interpreted, high-level, general-purpose programming language. Created by Guido Van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.



Logo of Python

Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released 2008, was a major revision of the language that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3. Due to concern about the amount of code written for Python 2, support for Python 2.7 (the last release in the 2.x series) was extended to 2020. Language developer Guido Van Rossum shouldered sole responsibility for the project until July 2018 but now shares his leadership as a member of a five-person steering council.

Python interpreters are available for many operating systems. A global community of programmers develops and maintains CPython, an open source reference implementation. A non-profit organization, the Python Software Foundation, manages and directs resources for Python and CPython development.

History → Python was conceived in the late 1980s by Guido Van Rossum at Centrum Wiskunde & Informatica (CWI) in the Netherlands as a successor to the ABC language (itself inspired by SETL), capable of exception handling and interfacing with the Amoeba operating system. Its implementation began in December 1989. Van Rossum continued as Python's lead developer until July 12, 2018, when he announced his "permanent vacation" from his responsibilities as Python's Benevolent Dictator for Life, a title the Python community bestowed upon him to reflect his long-term commitment as the project's chief decision-maker. In January, 2019, active Python core developers elected Brett Cannon, Nick Coghlan, Barry Warsaw, Carol Willing and Van Rossum to a five-member "Steering Council" to lead the project.

Guido Van Rossum (Creator of Python)

Python 2.0 was released on 16 October 2000 with many major new features, including a cycle-detecting garbage collector and support for Unicode.

Python 3.0 was released on 3 December 2008. It was a major revision of the language that is not completely backward-compatible. Many of its major features were backported to Python 2.6.x and 2.7.x version series. Releases of Python 3 include the 2to3 utility, which automates (at least partially) the translation of Python 2 code to Python 3.

Python 2.7's end-of-life date was initially set at 2015 then postponed to 2020 out of concern that a large body of existing code could not easily be forward-ported to Python 3. In January 2017, Google announced work on a Python 2.7 to Go transcompiler to improve performance under concurrent workloads.

Applications → Since 2003, Python has consistently ranked in the top ten most popular programming languages in the TIOBE Programming Community Index where, as of December 2018, it is the third most popular language (behind Java, and C). It was selected Programming Language of the Year in 2007, 2010, and 2018.

An empirical study found that scripting languages, such as Python, are more productive than conventional languages, such as C and Java, for programming problems involving string manipulation and search in a dictionary, and determined that memory consumption was often "better than Java and not much worse than C or C++".

Large organizations that use Python include Wikipedia, Google, Yahoo!, CERN, NASA, Facebook, Amazon, Instagram, Spotify and some smaller entities like ILM and ITA. The social news networking site Reddit is written entirely in Python.

Python can serve as a scripting language for web applications, e.g., via mod_wsgi for the Apache web server. With Web Server Gateway Interface, a standard API has evolved to facilitate these applications. Web frameworks like Django, Pylons, Pyramid, TurboGears, Web2py, Tornado, Flask, Bottle and Zope support developers in the design and maintenance of complex applications. Pyjs and IronPython can be used to develop the client-side of Ajax-based applications. SQLAlchemy can be used as data mapper to a relational database. Twisted is a framework to program communications between computers, and is used (for example) by Dropbox.

Libraries such as NumPy, SciPy and Matplotlib allow the effective use of Python in scientific computing, with specialized libraries such as Biopython and Astropy providing domain-specific functionality. SageMath is mathematical software with a notebook interface programmable in Python: its library covers many aspects of mathematics, including algebra, combinatorics, numerical mathematics, number theory, and calculus.

Python has been successfully embedded in many software products as a scripting language, including in finite element method software such as Abaqus, 3D parametric modeler like FreeCAD, 3D animation packages such as 3ds Max, Blender, Cinema 4D, Lightwave, Houdini, Maya, Modo, MotionBuilder, Softimage, the visual effects compositor Nuke, 2D imaging programs like GIMP, Inkscape, Scribus and Paint Shop Pro, and musical notation programs like Scorewriter and Capella. GNU Debugger uses Python as a pretty printer to show complex structures such as C++ containers. Esri promotes Python as the best choice for writing scripts in ArcGIS. It has also been used in several video games, and has been adopted as first of the three available programming languages in Google App Engine, the other two being Java and Go.

Python is commonly used in artificial intelligence projects with the help of libraries like TensorFlow, Keras and Scikit-learn. As a scripting language with modular architecture, simple syntax and rich text processing tools, Python is often used for natural language processing.

Many operating systems include Python as a standard component. It ships with most Linux distributions, AmigaOS 4, FreeBSD (as a package), NetBSD, OpenBSD (as a package) and MacOS and can be used from the command line (terminal). Many Linux distributions use

installers written in Python: Ubuntu uses the Ubiquity installer, while Red Hat Linux and Fedora use the Anaconda installer. Gentoo Linux uses Python in its package management system, Portage.

Python is used extensively in the information security industry, including in exploit development.

Most of the Sugar software for the One Laptop per Child XO, now developed at Sugar Labs, is written in Python. The Raspberry Pi single-board computer project has adopted Python as its main user-programming language.

LibreOffice includes Python, and intends to replace Java with Python. Its Python Scripting Provider is a core feature since Version 4.0 from 7 February 2013.

Naming → Python's name is derived from the British comedy group Monty Python, whom Python creator Guido Van Rossum enjoyed while developing the language. Monty Python references appear frequently in Python code and culture; for example, the metasyntactic variables often used in Python literature are spam and eggs instead of the traditional foo and bar. The official Python documentation also contains various references to Monty Python routines.

The prefix Py- is used to show that something is related to Python. Examples of the use of this prefix in names of Python applications or libraries include Pygame, a binding of SDL to Python (commonly used to create games); PyQt and PyGTK, which bind Qt and GTK to Python respectively; and PyPy, a Python implementation originally written in Python.

Artificial Intelligence

Introduction→ In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".

As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect. A quip in Tesler's Theorem says "AI is whatever hasn't been done yet." For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology. Modern machine capabilities generally classified as AI include successfully understanding human speech, competing at the highest level in strategic game systems (such as chess and Go), autonomously operating cars, intelligent routing in content delivery networks, and military simulations.

Artificial intelligence can be classified into three different types of systems: analytical, human-inspired, and humanized artificial intelligence. Analytical AI has only characteristics consistent with cognitive intelligence; generating cognitive representation of the world and using learning based on past experience to inform future decisions. Human-inspired AI has elements from cognitive and emotional intelligence; understanding human emotions, in addition to cognitive elements, and considering them in their decision making. Humanized AI shows characteristics of all types of competencies (i.e., cognitive, emotional, and social intelligence), is able to be self-conscious and is self-aware in interactions with others.

Artificial intelligence was founded as an academic discipline in 1956, and in the years since has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success and renewed funding. For most of its history, AI research has been divided into subfields that often fail to communicate with each other. These sub-fields are based on technical considerations, such as particular goals (e.g. "robotics" or "machine learning"), the use of particular tools ("logic" or artificial neural networks), or deep philosophical differences. Subfields have also been based on social factors (particular institutions or the work of particular researchers).

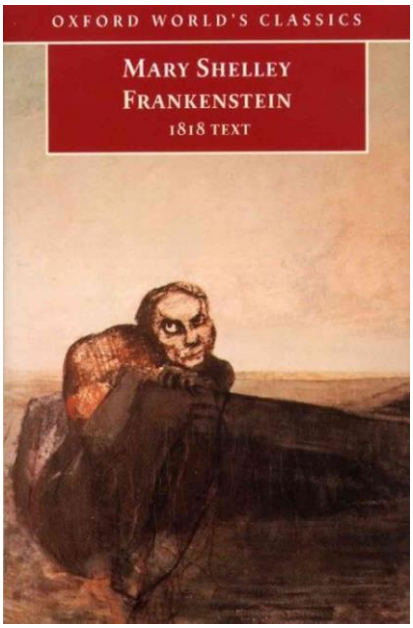
The traditional problems (or goals) of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception and the ability to move and manipulate objects. General intelligence is among the field's long-term goals. Approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, artificial neural networks, and methods based on statistics, probability and economics. The AI field draws upon

computer science, information engineering, mathematics, psychology, linguistics, philosophy, and many other fields.

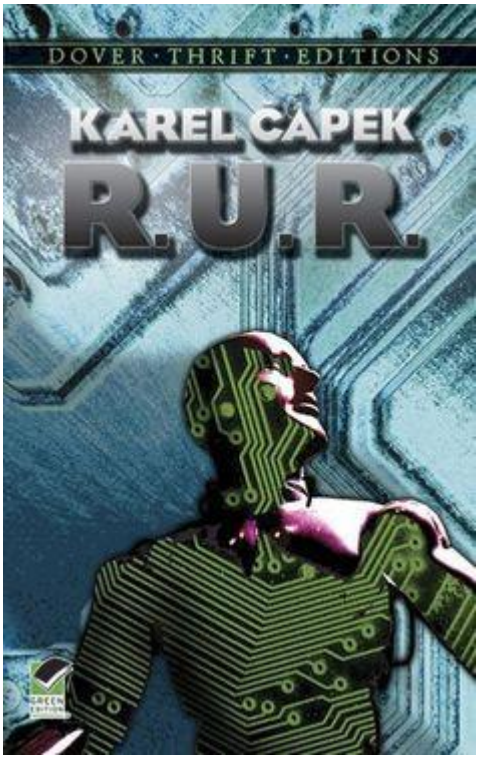
The field was founded on the claim that human intelligence "can be so precisely described that a machine can be made to simulate it". This raises philosophical arguments about the nature of the mind and the ethics of creating artificial beings endowed with human-like intelligence which are issues that have been explored by myth, fiction and philosophy since antiquity. Some people also consider AI to be a danger to humanity if it progresses unabated. Others believe that AI, unlike previous technological revolutions, will create a risk of mass unemployment.

In the twenty-first century, AI techniques have experienced a resurgence following concurrent advances in computer power, large amounts of data, and theoretical understanding; and AI techniques have become an essential part of the technology industry, helping to solve many challenging problems in computer science, software engineering and operations research.

History→ Thought-capable artificial beings appeared as storytelling devices in antiquity, and have been common in fiction, as in Mary Shelley's Frankenstein or Karel Capek's R.U.R. (Rossum's Universal Robots). These characters and their fates raised many of the same issues now discussed in the ethics of artificial intelligence.



Cover Page of the Book "Frankenstein" by writer Mary Shelley



Cover Page of the Book "Rossum's Universal Robots" by writer Karel Capek

The study of mechanical or "formal" reasoning began with philosophers and mathematicians in antiquity. The study of mathematical logic led directly to Alan Turing's theory of computation, which suggested that a machine, by shuffling symbols as simple as "0" and "1", could simulate any conceivable act of mathematical deduction. This insight, that digital computers can simulate any process of formal reasoning, is known as the Church–Turing thesis. Along with concurrent discoveries in neurobiology, information theory and cybernetics, this led researchers to consider the possibility of building an electronic brain. Turing proposed that "if a human could not distinguish between responses from a machine and a human, the machine could be considered intelligent". The first work that is now generally recognized as AI was McCullouch and Pitts' 1943 formal design for Turing-complete "artificial neurons".

The field of AI research was born at a workshop at Dartmouth College in 1956. Attendees Allen Newell (CMU), Herbert Simon (CMU), John McCarthy (MIT), Marvin Minsky (MIT) and Arthur Samuel (IBM) became the founders and leaders of AI research. They and their students produced programs that the press described as "astonishing" computers were learning checkers strategies (c. 1954) (and by 1959 were reportedly playing better than the average human),solving word problems in algebra, proving logical theorems and speaking English. By the middle of the 1960s, research in the U.S. was heavily funded by the Department of Defense and laboratories had been established around the world. AI's founders were optimistic about the future: Herbert Simon predicted, "Machines will be capable, within twenty years, of doing any work a man can do". Marvin Minsky agreed, writing, "Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved".

They failed to recognize the difficulty of some of the remaining tasks. Progress slowed and in 1974, in response to the criticism of Sir James Lighthill and ongoing pressure from the US Congress to fund more productive projects, both the U.S. and British governments cut off exploratory research in AI. The next few years would later be called an "AI winter", a period when obtaining funding for AI projects was difficult.

In the early 1980s, AI research was revived by the commercial success of expert systems, a form of AI program that simulated the knowledge and analytical skills of human experts. By 1985, the market for AI had reached over a billion dollars. At the same time, Japan's fifth generation computer project inspired the U.S and British governments to restore funding for academic research. However, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute, and a second, longer-lasting hiatus began.

In the late 1990s and early 21st century, AI began to be used for logistics, data mining, medical diagnosis and other areas. The success was due to increasing computational power (see Moore's law), greater emphasis on solving specific problems, new ties between AI and other fields (such as statistics, economics and mathematics), and a commitment by researchers to mathematical methods and scientific standards. Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov, on 11 May 1997.

In 2011, a Jeopardy! Quiz show exhibition match, IBM's question answering system, Watson, defeated the two greatest Jeopardy! Champions, Brad Rutter and Ken Jennings, by a significant margin. Faster computers, algorithmic improvements, and access to large amounts of data enabled advances in machine learning and perception; data-hungry deep learning methods started to dominate accuracy benchmarks around 2012. The Kinect, which provides a 3D body–motion interface for the Xbox 360 and the Xbox One, uses algorithms that emerged from lengthy AI research as do intelligent personal assistants in smartphones. In March 2016, AlphaGo won 4 out of 5 games of Go in a match with Go champion Lee Sedol, becoming the first computer Go-playing system to beat a professional Go player without handicaps. In the 2017 Future of Go Summit, AlphaGo won a three-game match with Ke Jie, who at the time continuously held the world No. 1 ranking for two years. This marked the completion of a significant milestone in the development of Artificial Intelligence as Go is a relatively complex game, more so than Chess.

According to Bloomberg's Jack Clark, 2015 was a landmark year for artificial intelligence, with the number of software projects that use AI Google increased from a "sporadic usage" in 2012 to more than 2,700 projects. Clark also presents factual data indicating the improvements of AI since 2012 supported by lower error rates in image processing tasks. He attributes this to an increase in affordable neural networks, due to a rise in cloud computing infrastructure and to an increase in research tools and datasets. Other cited examples include Microsoft's development of a Skype system that can automatically translate from one language to another and Facebook's system that can describe images to blind people. In a 2017 survey, one in five companies reported they had "incorporated AI in some offerings or processes". Around 2016, China greatly accelerated its government funding; given its large supply of data and its rapidly increasing research output, some observers believe it May be on track to becoming an "AI superpower".

However, it has been acknowledged that reports regarding artificial intelligence have tended to be exaggerated.

Applications→ AI is relevant to any intellectual task. Modern artificial intelligence techniques are pervasive and are too numerous to list here. Frequently, when a technique reaches mainstream use, it is no longer considered artificial intelligence; this phenomenon is described as the AI effect.

High-profile examples of AI include autonomous vehicles (such as drones and self-driving cars), medical diagnosis, creating art (such as poetry), proving mathematical theorems, playing games (such as Chess or Go), search engines (such as Google search), online assistants (such as Siri), image recognition in photographs, spam filtering, predicting flight delays, prediction of judicial decisions and targeting online advertisements.

With social media sites overtaking TV as a source for news for young people and news organizations increasingly reliant on social media platforms for generating distribution, major publishers now use artificial intelligence (AI) technology to post stories more effectively and generate higher volumes of traffic.

1) **Healthcare :**

AI is being applied to the high cost problem of dosage issues—where findings suggested that AI could save $16 billion. In 2016, a ground breaking study in California found that a mathematical formula developed with the help of AI correctly determined the accurate dose of immunosuppressant drugs to give to organ patients.

Artificial intelligence is breaking into the healthcare industry by assisting doctors. According to Bloomberg Technology, Microsoft has developed AI to help doctors find the right treatments for cancer. There is a great amount of research and drugs developed relating to cancer. In detail, there are more than 800 medicines and vaccines to treat cancer. This negatively affects the doctors, because there are too many options to choose from, making it more difficult to choose the right drugs for the patients. Microsoft is working on a project to develop a machine called "Hanover". Its goal is to memorize all the papers necessary to cancer and help predict which combinations of drugs will be most effective for each patient. One project that is being worked on at the moment is fighting myeloid leukemia, a fatal cancer where the treatment has not improved in decades. Another study was reported to have found that artificial intelligence was as good as trained doctors in identifying skin cancers. Another study is using artificial intelligence to try and monitor multiple high-risk patients, and this is done by asking each patient numerous questions based on data acquired from live doctor to patient interactions. One study was done with transfer learning; the machine performed a diagnosis similarly to a well-trained ophthalmologist, and could generate a decision within 30 seconds on whether or not the patient should be referred for treatment, with more than 95% accuracy.

According to CNN, a recent study by surgeons at the Children's National Medical Center in Washington successfully demonstrated surgery with an autonomous robot. The team supervised the robot while it performed soft-tissue surgery, stitching together a pig's bowel during open

surgery, and doing so better than a human surgeon, the team claimed. IBM has created its own artificial intelligence computer, the IBM Watson, which has beaten human intelligence (at some levels). Watson not only won at the game show Jeopardy! Against former champions, but was declared a hero after successfully diagnosing a woman who was suffering from leukemia.

2) Automotive :

Advancements in AI have contributed to the growth of the automotive industry through the creation and evolution of self-driving vehicles. As of 2016, there are over 30 companies utilizing AI into the creation of driverless cars. A few companies involved with AI include Tesla, Google, and Apple.

Many components contribute to the functioning of self-driving cars. These vehicles incorporate systems such as braking, lane changing, collision prevention, navigation and mapping. Together, these systems, as well as high performance computers, are integrated into one complex vehicle.

Recent developments in autonomous automobiles have made the innovation of self-driving trucks possible, though they are still in the testing phase. The UK government has passed legislation to begin testing of self-driving truck platoons in 2018. Self-driving truck platoons are a fleet of self-driving trucks following the lead of one non-self-driving truck, so the truck platoons aren't entirely autonomous yet. Meanwhile, the Daimler, a German automobile corporation, is testing the Freightliner Inspiration which is a semi-autonomous truck that will only be used on the highway.

One main factor that influences the ability for a driver-less automobile to function is mapping. In general, the vehicle would be pre-programmed with a map of the area being driven. This map would include data on the approximations of street light and curb heights in order for the vehicle to be aware of its surroundings. However, Google has been working on an algorithm with the purpose of eliminating the need for pre-programmed maps and instead, creating a device that would be able to adjust to a variety of new surroundings. Some self-driving cars are not equipped with steering wheels or brake pedals, so there has also been research focused on creating an algorithm that is capable of maintaining a safe environment for the passengers in the vehicle through awareness of speed and driving conditions.

Another factor that is influencing the ability for a driver-less automobile is the safety of the passenger. To make a driver-less automobile, engineers must program it to handle high-risk situations. These situations could include a head-on collision with pedestrians. The car's main goal should be to make a decision that would avoid hitting the pedestrians and saving the passengers in the car. But there is a possibility the car would need to make a decision that would put someone in danger. In other words, the car would need to decide to save the pedestrians or the passengers. The programming of the car in these situations is crucial to a successful driver-less automobile.

3) Finance and Economics :

Financial institutions have long used artificial neural network systems to detect charges or claims outside of the norm, flagging these for human investigation. The use of AI in banking can be traced back to 1987 when Security Pacific National Bank in US set-up a Fraud Prevention Task force to counter the unauthorized use of debit cards. Programs like Kasisto and Moneystream are using AI in financial services.

Banks use artificial intelligence systems today to organize operations, maintain book-keeping, invest in stocks, and manage properties. AI can react to changes overnight or when business is not taking place. In August 2001, robots beat humans in a simulated financial trading competition. AI has also reduced fraud and financial crimes by monitoring behavioral patterns of users for any abnormal changes or anomalies.

The use of AI machines in the market in applications such as online trading and decision making has changed major economic theories. For example, AI based buying and selling platforms have changed the law of supply and demand in that it is now possible to easily estimate individualized demand and supply curves and thus individualized pricing. Furthermore, AI machines reduce information asymmetry in the market and thus making markets more efficient while reducing the volume of trades. Furthermore, AI in the markets limits the consequences of behavior in the markets again making markets more efficient. Other theories where AI has had impact include in rational choice, rational expectations, game theory, Lewis turning point, portfolio optimization and counterfactual thinking.

4) Government :

Artificial intelligence paired with facial recognition systems May be used for mass surveillance. This is already the case in some parts of China.

5) Video games :

In video games, artificial intelligence is routinely used to generate dynamic purposeful behavior in non-player characters (NPCs). In addition, well-understood AI techniques are routinely used for pathfinding. Some researchers consider NPC AI in games to be a "solved problem" for most production tasks. Games with more atypical AI include the AI director of Left 4 Dead (2008) and the evolutionary training of platoons in Supreme Commander 2 (2010).

6) Military :

Worldwide annual military spending on robotics rose from US$5.1 billion in 2010 to US$7.5 billion in 2015. Military drones capable of autonomous action are widely considered a useful asset. Many artificial intelligence researchers seek to distance themselves from military applications of AI.

7) Audit :

For financial statements audit, AI makes continuous audit possible. AI tools could analyze many sets of different information immediately. The potential benefit would be the overall audit

risk will be reduced, the level of assurance will be increased and the time duration of audit will be reduced.

8) Advertising :

It is possible to use AI to predict or generalize the behavior of customers from their digital footprints in order to target them with personalized promotions or build customer personas automatically. A documented case reports that online gambling companies were using AI to improve customer targeting.

Moreover, the application of Personality computing AI models can help reducing the cost of advertising campaigns by adding psychological targeting to more traditional socio-demographic or behavioral targeting.

9) Art :

Artificial Intelligence has inspired numerous creative applications including its usage to produce visual art. The exhibition "Thinking Machines: Art and Design in the Computer Age, 1959–1989" at MoMA provides a good overview of the historical applications of AI for art, architecture, and design. Recent exhibitions showcasing the usage of AI to produce art include the Google-sponsored benefit and auction at the Gray Area Foundation in San Francisco, where artists experimented with the deep-dream algorithm and the exhibition "Unhuman: Art in the Age of AI," which took place in Los Angeles and Frankfurt in the fall of 2017. In the spring of 2018, the Association of Computing Machinery dedicated a special magazine issue to the subject of computers and art highlighting the role of machine learning in the arts.

Objectives

An Ecommerce Company Snapdeal has many customers across the world. Customers buy products from Snapdeal either by Website or by Smartphone App. Now the company is thinking to decide whether to focus their efforts on either Smartphone App based Sale or Website based Sale, so that their profit is maximum.

With the help of Python and Artificial Intelligence, this innovative project analyses the sales data of Snapdeal Company, and gives an optimized model for the Snapdeal company to take the decision.

Hardware and Softwares

Hardware

Machine → HP BS658TX (Laptop)

RAM Memory → 8 GB

Hard Disk Memory → 1024 GB

Softwares

Operating System → Windows 10 Enterprise (64 bit)

Application Software → Anaconda 3

Programming Language → Python 3.7.3

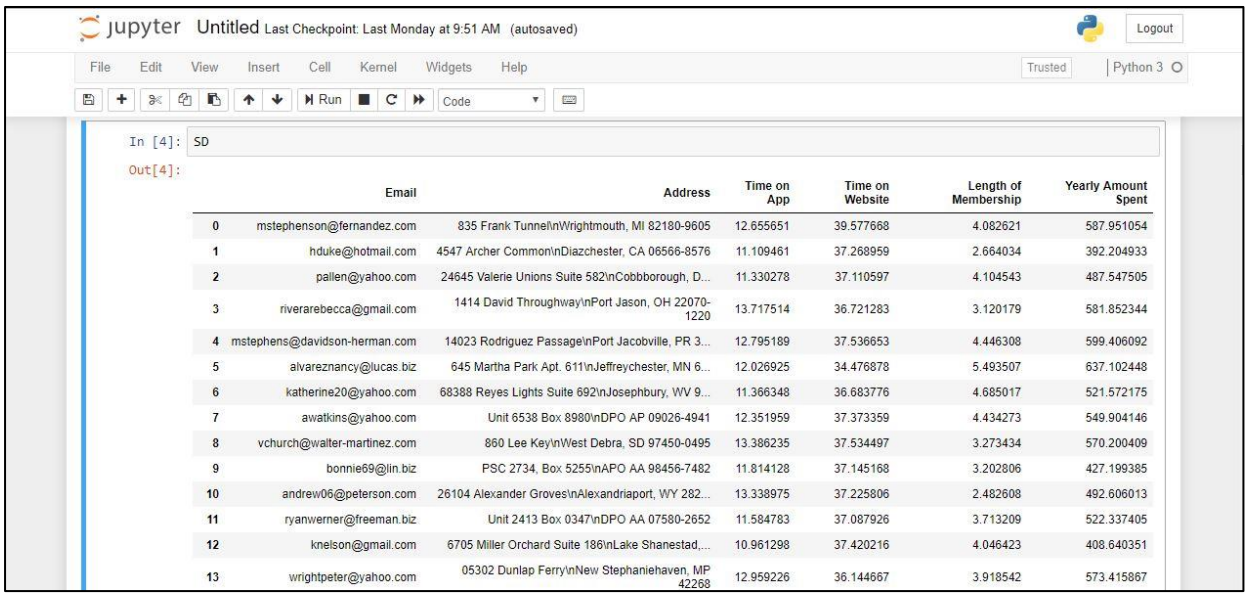
Code and Output

(Codes are given in black ink. Comments about codes are given in Green ink. Output is given to corresponding input, as a screenshot.)

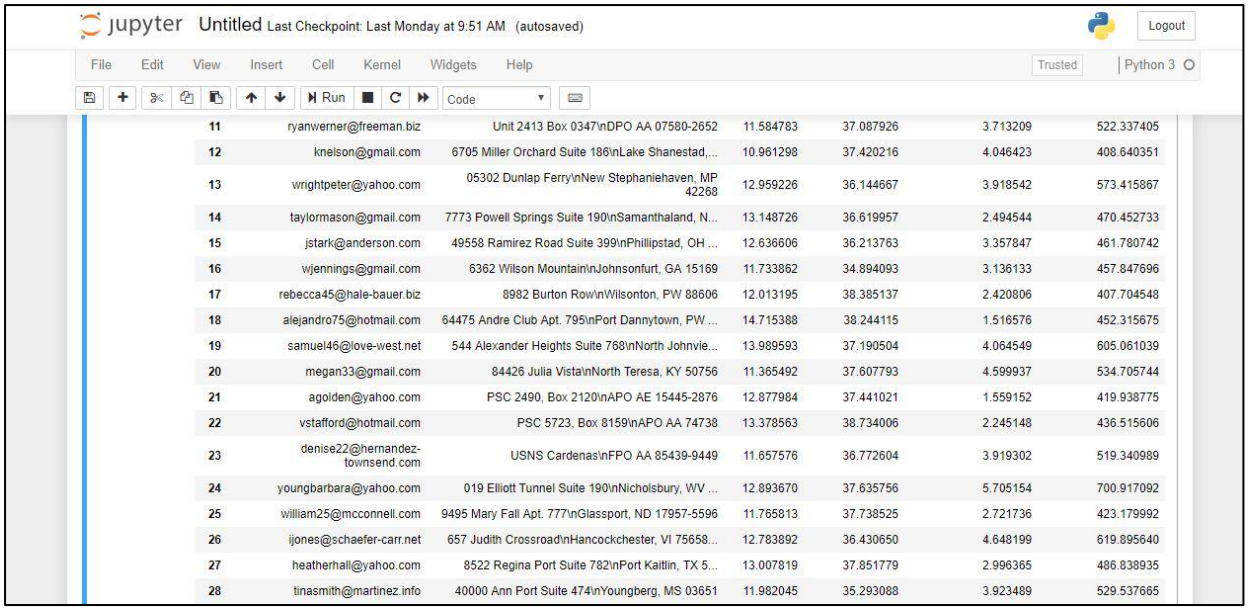
```
import pandas as pd      ←(Import Pandas)
import seaborn as sns    ←(Import Seaborn)
import numpy as np       ←(Import Numpy)
import matplotlib.pyplot as plt    ←(Import Matplotlib)
sns.set_style('whitegrid')    ←(Set Grid Style as White for all Graphs)
%matplotlib inline
```

```
SD = pd.read_csv('Snapdeal Dataset.csv')    ←(Read the sales data of Snapdeal Company)
```

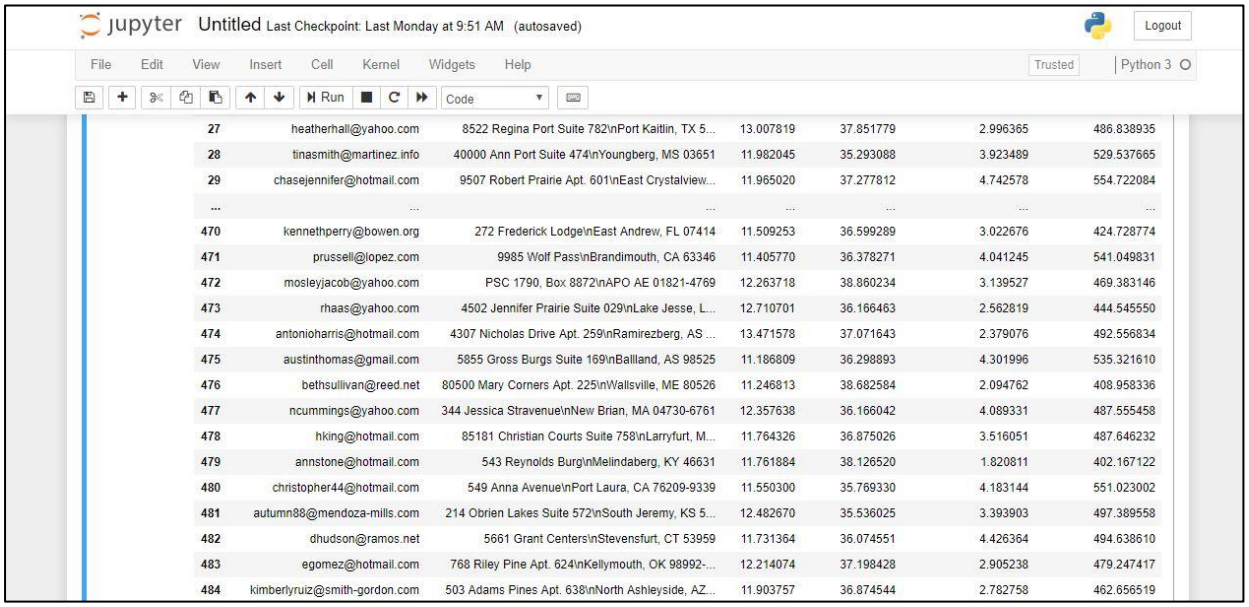
```
SD    ←(Show the SD Dataframe)
```



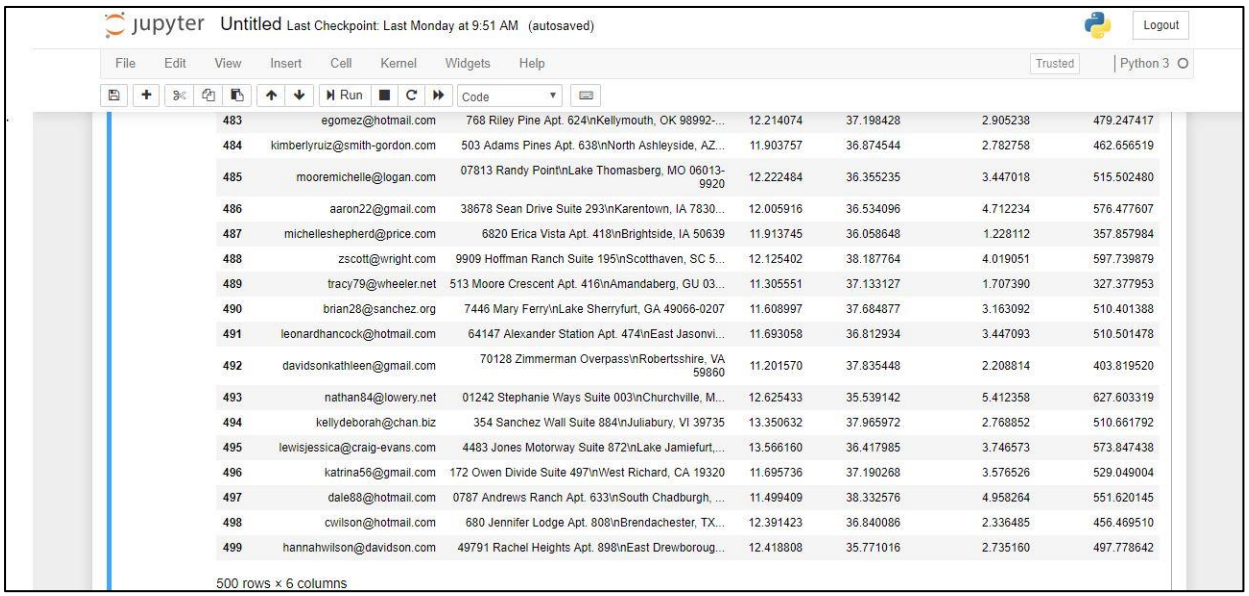
	Email	Address	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
0	mstephenson@fernandez.com	835 Frank Tunnel\nWrightmouth, MI 82180-9605	12.655651	39.577668	4.082621	587.951054
1	hduke@hotmail.com	4547 Archer Common\nDiazchester, CA 06566-8576	11.109461	37.268959	2.664034	392.204933
2	pallen@yahoo.com	24645 Valerie Unions Suite 582\nCobbborough, D...	11.330278	37.110597	4.104543	487.547505
3	riverarebecca@gmail.com	1414 David Throughway\nPort Jason, OH 22070-1220	13.717514	36.721283	3.120179	581.852344
4	mstephens@davidson-herman.com	14023 Rodriguez Passage\nPort Jacobville, PR 3...	12.795189	37.536653	4.446308	599.406092
5	alvareznancy@lucas.biz	645 Martha Park Apt. 611\nJeffreychester, MN 6...	12.026925	34.476878	5.493507	637.102448
6	katherine20@yahoo.com	68388 Reyes Lights Suite 692\nJosephbury, WV 9...	11.366348	36.683776	4.685017	521.572175
7	awatkins@yahoo.com	Unit 6538 Box 8980\nDPO AP 09026-4941	12.351959	37.373359	4.434273	549.904146
8	vchurch@walter-martinez.com	860 Lee Key\nWest Debra, SD 97450-0495	13.386235	37.534497	3.273434	570.200409
9	bonnie69@lin.biz	PSC 2734, Box 5255\nAPO AA 98456-7482	11.814128	37.145168	3.202806	427.199385
10	andrew06@peterson.com	26104 Alexander Groves\nAlexandriaport, WY 282...	13.338975	37.225806	2.482608	492.606013
11	ryanwerner@freeman.biz	Unit 2413 Box 0347\nDPO AA 07580-2652	11.584783	37.087926	3.713209	522.337405
12	knelson@gmail.com	6705 Miller Orchard Suite 186\nLake Shanestad,...	10.961298	37.420216	4.046423	408.640351
13	wrightpeter@yahoo.com	05302 Dunlap Ferry\nNew Stephaniehaven, MP 42268	12.959226	36.144667	3.918542	573.415867



11	ryanwerner@freeman.biz	Unit 2413 Box 0347\nDPO AA 07580-2652	11.584783	37.087926	3.713209	522.337405
12	knelson@gmail.com	6705 Miller Orchard Suite 186\nLake Shanestad,...	10.961298	37.420216	4.046423	408.640351
13	wrightpeter@yahoo.com	05302 Dunlap Ferry\nNew Stephaniehaven, MP 42268	12.959226	36.144667	3.918542	573.415867
14	taylormason@gmail.com	7773 Powell Springs Suite 190\nSamanthaland, N...	13.148726	36.619957	2.494544	470.452733
15	jstark@anderson.com	49558 Ramirez Road Suite 399\nPhillipstad, OH ...	12.636606	36.213763	3.357847	461.780742
16	wjennings@gmail.com	6362 Wilson Mountain\nJohnsonfurt, GA 15169	11.733862	34.894093	3.136133	457.847696
17	rebecca45@hale-bauer.biz	8982 Burton Row\nWilsonton, PW 88606	12.013195	38.385137	2.420806	407.704548
18	alejandro75@hotmail.com	64475 Andre Club Apt. 795\nPort Dannytown, PW ...	14.715388	38.244115	1.516576	452.315675
19	samuel46@love-west.net	544 Alexander Heights Suite 768\nNorth Johnvie...	13.989593	37.190504	4.064549	605.061039
20	megan33@gmail.com	84426 Julia Vista\nNorth Teresa, KY 50756	11.365492	37.607793	4.599937	534.705744
21	agolden@yahoo.com	PSC 2490, Box 2120\nAPO AE 15445-2876	12.877984	37.441021	1.559152	419.938775
22	vstafford@hotmail.com	PSC 5723, Box 8159\nAPO AA 74738	13.378563	38.734006	2.245148	436.515606
23	denise22@hernandez-townsend.com	USNS Cardenas\nFPO AA 85439-9449	11.657576	36.772604	3.919302	519.340989
24	youngbarbara@yahoo.com	019 Elliott Tunnel Suite 190\nNicholsbury, WV ...	12.893670	37.635756	5.705154	700.917092
25	william25@mcconnell.com	9495 Mary Fall Apt. 777\nGlassport, ND 17957-5596	11.765813	37.738525	2.721736	423.179992
26	ijones@schaefer-carr.net	657 Judith Crossroad\nHancockchester, VI 75658...	12.783892	36.430650	4.648199	619.895640
27	heatherhall@yahoo.com	8522 Regina Port Suite 782\nPort Kaitlin, TX 5...	13.007819	37.851779	2.996365	486.838935
28	tinasmith@martinez.info	40000 Ann Port Suite 474\nYoungberg, MS 03651	11.982045	35.293088	3.923489	529.537665



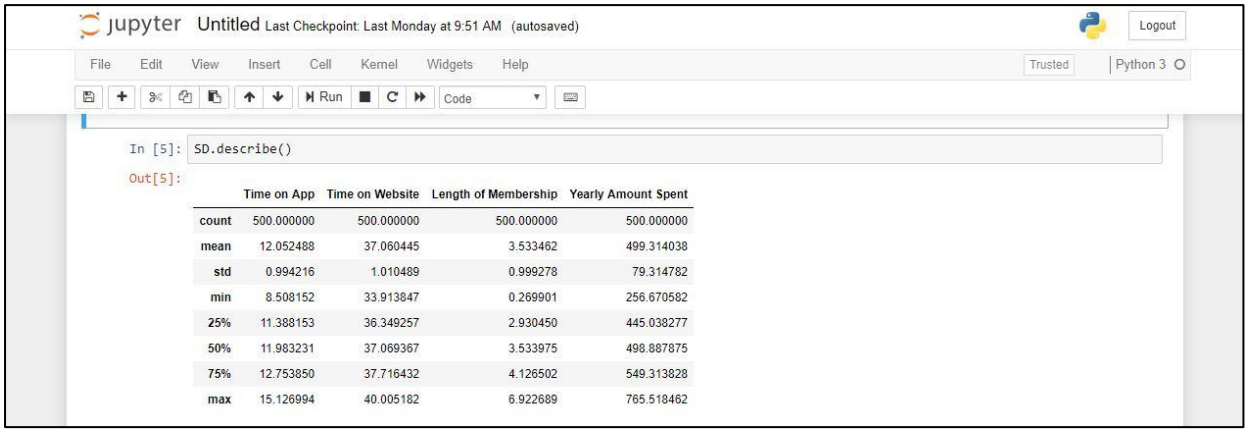
27	heatherhall@yahoo.com	8522 Regina Port Suite 782\nPort Kaitlin, TX 5...	13.007819	37.851779	2.996365	486.838935
28	tinasmith@martinez.info	40000 Ann Port Suite 474\nYoungberg, MS 03651	11.982045	35.293088	3.923489	529.537665
29	chasejennifer@hotmail.com	9507 Robert Prairie Apt. 601\nEast Crystalview...	11.965020	37.277812	4.742578	554.722084
...	...	...	...	...	...	...
470	kennethperry@bowen.org	272 Frederick Lodge\nEast Andrew, FL 07414	11.509253	36.599289	3.022676	424.728774
471	prussell@lopez.com	9985 Wolf Pass\nBrandimouth, CA 63346	11.405770	36.378271	4.041245	541.049831
472	mosleyjacob@yahoo.com	PSC 1790, Box 8872\nAPO AE 01821-4769	12.263718	38.860234	3.139527	469.383146
473	rhaas@yahoo.com	4502 Jennifer Prairie Suite 029\nLake Jesse, L...	12.710701	36.166463	2.562819	444.545550
474	antonioharris@hotmail.com	4307 Nicholas Drive Apt. 259\nRamirezberg, AS ...	13.471578	37.071643	2.379076	492.556834
475	austinthomas@gmail.com	5855 Gross Burgs Suite 169\nBallland, AS 98525	11.186809	36.298893	4.301996	535.321610
476	bethsullivan@reed.net	80500 Mary Corners Apt. 225\nWallsville, ME 80526	11.246813	38.682584	2.094762	408.958336
477	ncummings@yahoo.com	344 Jessica Stravenue\nNew Brian, MA 04730-6761	12.357638	36.166042	4.089331	487.555458
478	hking@hotmail.com	85181 Christian Courts Suite 758\nLarryfurt, M...	11.764326	36.875026	3.516051	487.646232
479	annstone@hotmail.com	543 Reynolds Burg\nMelindaberg, KY 46631	11.761884	38.126520	1.820811	402.167122
480	christopher44@hotmail.com	549 Anna Avenue\nPort Laura, CA 76209-9339	11.550300	35.769330	4.183144	551.023002
481	autumn88@mendoza-mills.com	214 Obrien Lakes Suite 572\nSouth Jeremy, KS 5...	12.482670	35.536025	3.393903	497.389558
482	dhudson@ramos.net	5661 Grant Centers\nStevensfurt, CT 53959	11.731364	36.074551	4.426364	494.638610
483	egomez@hotmail.com	768 Riley Pine Apt. 624\nKellymouth, OK 98992-...	12.214074	37.198428	2.905238	479.247417
484	kimberlyruiz@smith-gordon.com	503 Adams Pines Apt. 638\nNorth Ashleyside, AZ...	11.903757	36.874544	2.782758	462.656519



483	egomez@hotmail.com	768 Riley Pine Apt. 624\nKellymouth, OK 98992-...	12.214074	37.198428	2.905238	479.247417
484	kimberlyruiz@smith-gordon.com	503 Adams Pines Apt. 638\nNorth Ashleyside, AZ...	11.903757	36.874544	2.782758	462.656519
485	mooremichelle@logan.com	07813 Randy Point\nLake Thomasberg, MO 06013-9920	12.222484	36.355235	3.447018	515.502480
486	aaron22@gmail.com	38678 Sean Drive Suite 293\nKarentown, IA 7830...	12.005916	36.534096	4.712234	576.477607
487	michelleshepherd@price.com	6820 Erica Vista Apt. 418\nBrightside, IA 50639	11.913745	36.058648	1.228112	357.857984
488	zscott@wright.com	9909 Hoffman Ranch Suite 195\nScotthaven, SC 5...	12.125402	38.187764	4.019051	597.739879
489	tracy79@wheeler.net	513 Moore Crescent Apt. 416\nAmandaberg, GU 03...	11.305551	37.133127	1.707390	327.377953
490	brian28@sanchez.org	7446 Mary Ferry\nLake Sherryfurt, GA 49066-0207	11.608997	37.684877	3.163092	510.401388
491	leonardhancock@hotmail.com	64147 Alexander Station Apt. 474\nEast Jasonvi...	11.693058	36.812934	3.447093	510.501478
492	davidsonkathleen@gmail.com	70128 Zimmerman Overpass\nRobertsshire, VA 59860	11.201570	37.835448	2.208814	403.819520
493	nathan84@lowery.net	01242 Stephanie Ways Suite 003\nChurchville, M...	12.625433	35.539142	5.412358	627.603319
494	kellydeborah@chan.biz	354 Sanchez Wall Suite 884\nJuliabury, VI 39735	13.350632	37.965972	2.768852	510.661792
495	lewisjessica@craig-evans.com	4483 Jones Motorway Suite 872\nLake Jamiefurt,...	13.566160	36.417985	3.746573	573.847438
496	katrina56@gmail.com	172 Owen Divide Suite 497\nWest Richard, CA 19320	11.695736	37.190268	3.576526	529.049004
497	dale88@hotmail.com	0787 Andrews Ranch Apt. 633\nSouth Chadburgh, ...	11.499409	38.332576	4.958264	551.620145
498	cwilson@hotmail.com	680 Jennifer Lodge Apt. 808\nBrendachester, TX...	12.391423	36.840086	2.336485	456.469510
499	hannahwilson@davidson.com	49791 Rachel Heights Apt. 898\nEast Drewboroug...	12.418808	35.771016	2.735160	497.778642

500 rows × 6 columns

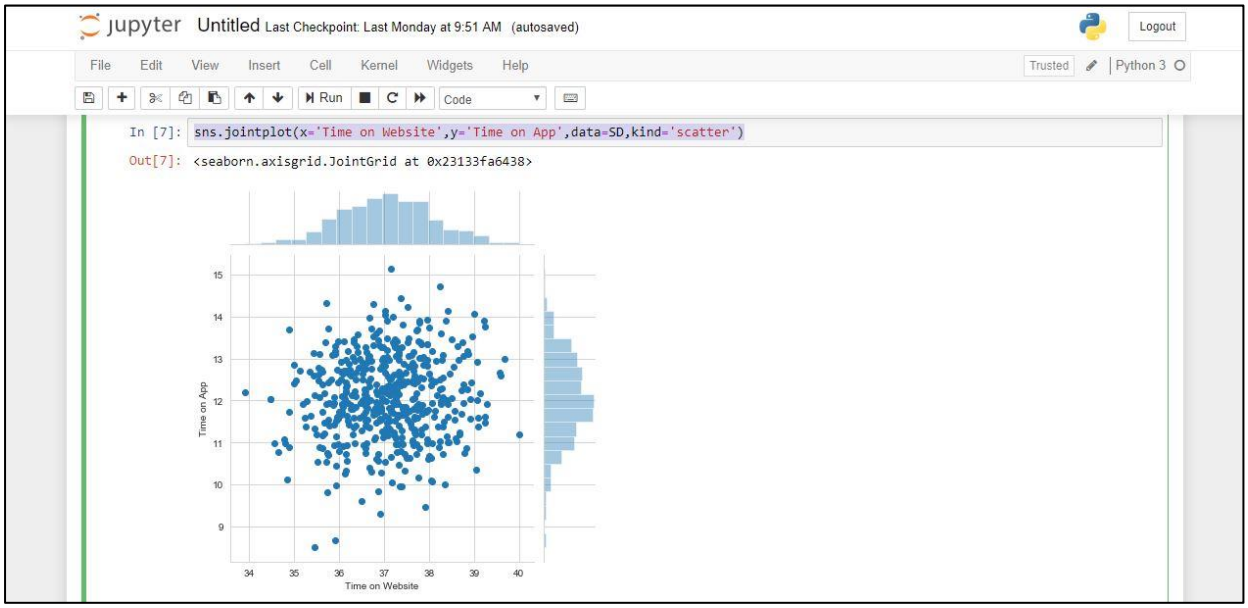
SD.describe( ) ←(Calculate total row wise entries of dataset. Then Calculate Average, Standard Deviation, Minimum value and Maximum value among Column Attributes)



	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
count	500.000000	500.000000	500.000000	500.000000
mean	12.052488	37.060445	3.533462	499.314038
std	0.994216	1.010489	0.999278	79.314782
min	8.508152	33.913847	0.269901	256.670582
25%	11.388153	36.349257	2.930450	445.038277
50%	11.983231	37.069367	3.533975	498.887875
75%	12.753850	37.716432	4.126502	549.313828
max	15.126994	40.005182	6.922689	765.518462

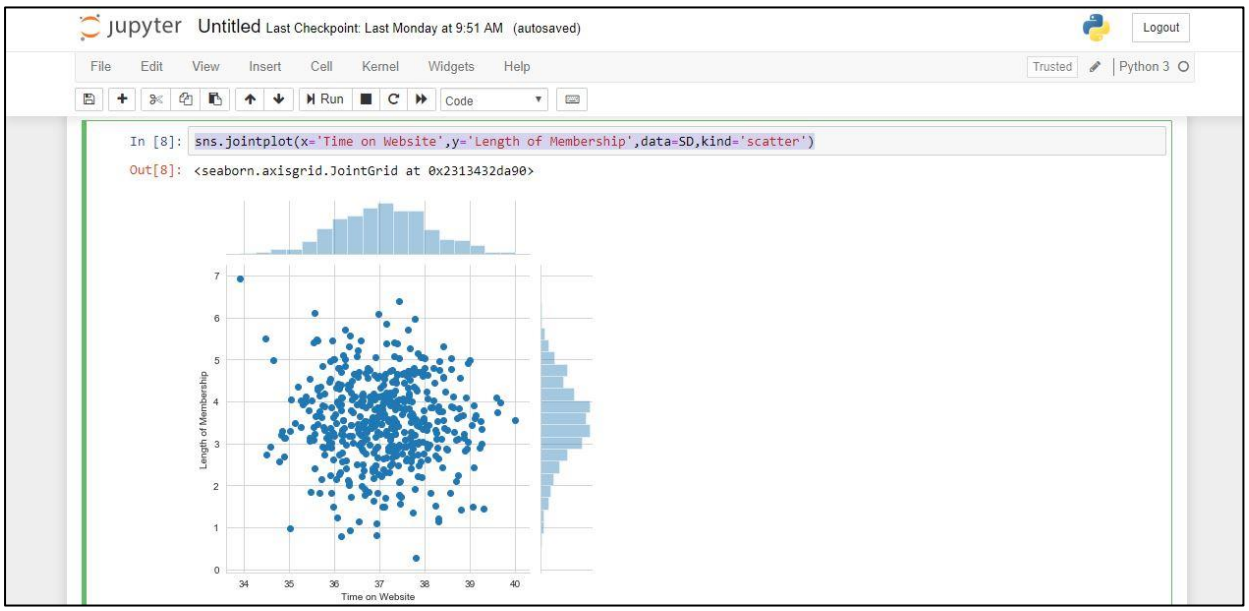
sns.jointplot(x = 'Time on Website', y = 'Time on App', data = SD, kind = 'scatter')

←(Create a 2D Cartesian graph among Time on Website, and Time on App)



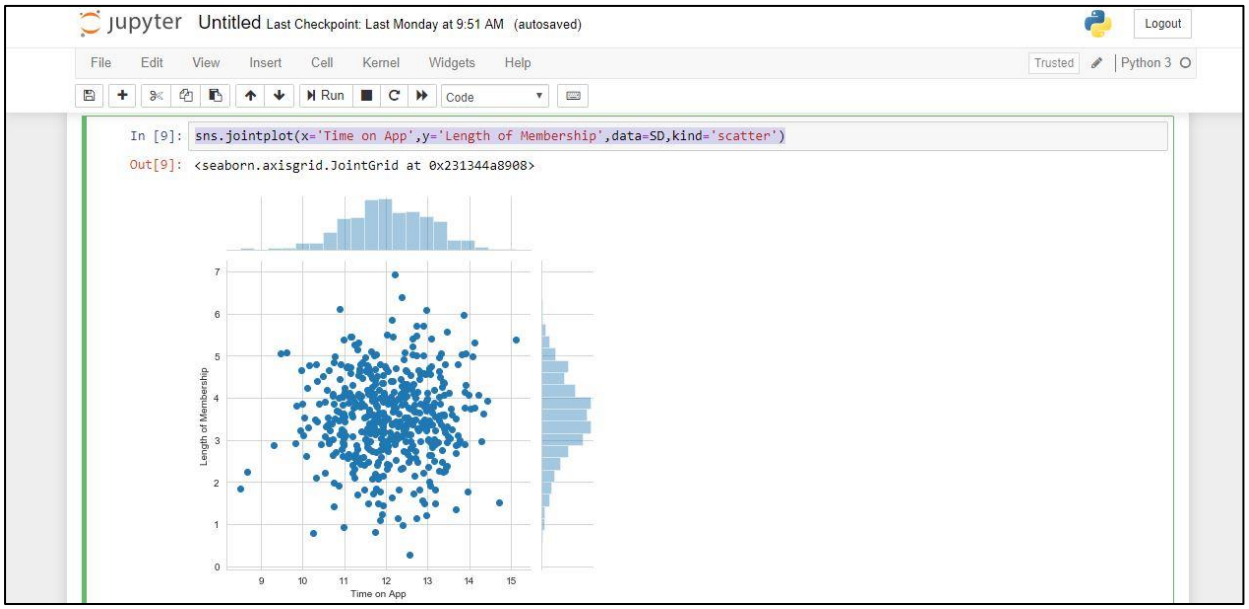
sns.jointplot(x = 'Time on Website', y = 'Length of Membership', data = SD, kind = 'scatter')

←(Create a 2D Cartesian graph among Time on Website, and Length of Membership)



sns.jointplot(x = 'Time on App', y = 'Length of Membership', data = SD, kind = 'scatter')

←(Create a 2D Cartesian graph among Time on App, and Length of Membership)



from sklearn.model_selection import train_test_split ←(Import train_test_split)

X=SD[['Time on App','Time on Website','Length of Membership']]

y=SD['Yearly Amount Spent']

←(Divide the dataset into X and y to train and test data)

X_train,X_test,y_train,y_test = train_test_split(X, y, test_size=0.5, random_state=200)

←(Define the variables, test size and random state for training and testing)

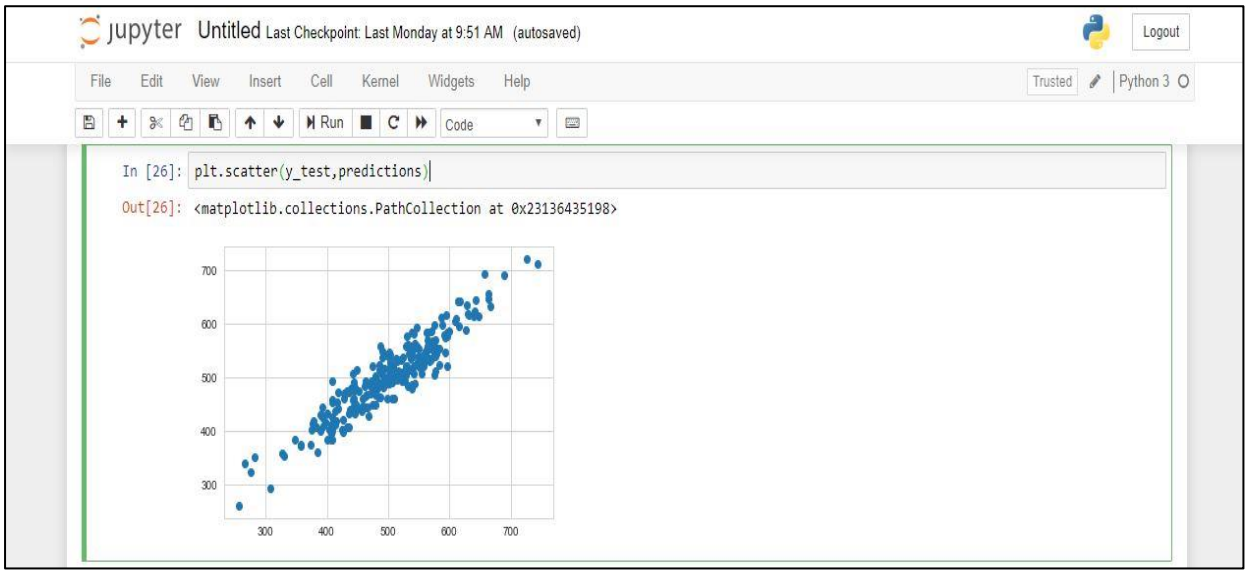
from sklearn.linear_model import LinearRegression ←(Import LinearRegression)

lm = LinearRegression() ←(Create an instance of LinearRegression( ))

lm.fit(X_train,y_train) ←(Fit the dataset for applying Linear Regression Concept)

predictions=lm.predict(X_test) ←(Store the predictive data of X_test in predictions variable)

plt.scatter(y_test,predictions) ←(Create 2D Cartesian graph of predictive model)



coeff_df = pd.DataFrame(lm.coef_,X.columns,columns=['Coefficient'])

←(Calculate Co efficients of Predictive Model)

coeff_df ←(Show the Co efficients of Predictive Model)

Discussion

In this project, at first, let us look on the average values of “Time on App”, “Time on Website”, “Length of Membership” and “Yearly amount spent”. The values are 12.05, 37.06, 3.53, and 499.31 simultaneously. So it seems that “Yearly amount spent” is more important, having highest average value. But more analysis of dataset is needed.

Then some 2D Cartesian graphs are plotted among “Time on App”, “Time on Website” and “Length of Membership”.

Then, concept of Artificial Intelligence is applied on dataset by splitting it into training set and testing set. Then a Predictive Model of dataset is built by concept of linear regression. Then a 2D Cartesian graph is plotted of Predictive Model.

All of these 2D graphs suggest that any one thing among “Time on App”, “Time on Website”, and “Length of Membership” needs to be considered.

To take the final decision, the co-efficient of “Time on App”, “Time on Website”, and “Length of Membership” are calculated. Now, since the co efficient of “Length of Membership” is highest, so that is considered to be most important.

Python programming language is used in this project; because it is very easy to draw graph and statistical analysis on a database, by python.

Making Predictive Model by training dataset and testing dataset, is a concept of artificial intelligence.

Linear regression is a concept of Statistics.

Conclusion

After analyzing the dataset, it is clear that Length of Membership is most important.

So, Snapdeal Company should concentrate on the Length of Membership of Snapdeal Users, to maximize their profit and to make best business strategies.

Bibliography

1) https://en.wikipedia.org/wiki/Python_%28programming_language%29

2) https://en.wikipedia.org/wiki/Artificial_intelligence