



The Ultimate Handbook for
**Building Data Analytics
Portfolio and Projects**





Table of Contents

▪ Introduction	02
▪ Significance of Having a Portfolio	03
▪ Advantages of Building a Portfolio	04
▪ Job Trends & Projections for Data Analysts	05
▪ Industry-Relevant Skills for Data Analysts	06
▪ What Should a Data Analytics Portfolio Demonstrate?	07
▪ What are Data Analytics Projects?	08
▪ Importance of Doing Data Analytics Projects	09
▪ Recommended Data Analytics Projects for Beginners	10
▪ Recommended Data Analysis Projects for Mid-Level Professionals	16
▪ Other Projects That You Can Try Your Hands On	22
▪ Must-Have Data Analysis Tools in Your Arsenal	23
▪ Additional Resources That Data Analysts Can Use	26
▪ How to Land a Job in the Field of Data Analytics?	28
▪ About DASCA	30

In today's digitally driven world, information and technology have become ubiquitous. One significant market that has emerged as a result of technological advancements is data analytics. It has now become an essential component of enterprises in numerous sectors, providing comprehensive market analysis and valuable insights into customer behavior, which can give businesses a competitive advantage.

Currently, there is no industry that does not gather, evaluate, and capture insights from data, which is why data analysis skills are versatile in the 21st century. Additionally, this has led to fierce competition for these positions. How can you distinguish yourself from the other applicants when everyone is enthusiastic, diligent, and knowledgeable? The answer to this is an analytics portfolio.

This guide provides valuable insights into the benefits of having a portfolio and offers a range of significant projects that can be included to help you get started or accelerate your career in data science.



Significance of Having a Portfolio

A well-built portfolio showcases your growth in understanding and skills over time, essentially extending beyond the information on your resume. While your resume lists your skills and achievements for potential employers, a portfolio demonstrates those avant-garde skills through documented accomplishments.

When creating an analytics portfolio, the objectives involve demonstrating technical proficiency to a recruiting manager. This is especially crucial if you are new to the industry, have a non-technical educational background, or are changing careers. To effectively convey outcomes, you will need to dedicate time to developing algorithms and implementing fixes.

A dynamic data analytics portfolio should include:



Projects

Working on projects exposes you to various scenarios, challenges, and opportunities that may not be available in a typical classroom setting. Furthermore, the skills and knowledge gained from projects can help you to stand out from other candidates during a job interview.



Internships

Internships are the closest thing you can have to a job. These opportunities give you a taste of a real job and shine the brightest on your portfolio as a testament to your ability to handle them.



Professional Certifications

The popularity of professional certifications is increasing among data science professionals. This is because certifications from reputable sources can land you your dream job as a data analyst, even if you are majoring in a subject other than data science.



Validation

A job requires more than just hard skills. Reviews, referrals, and testimonials from past colleagues, bosses, clients, and teachers work as a great reassurance for future employers. Your data analytics portfolio is incomplete without such social validation.

Advantages of Building a Portfolio

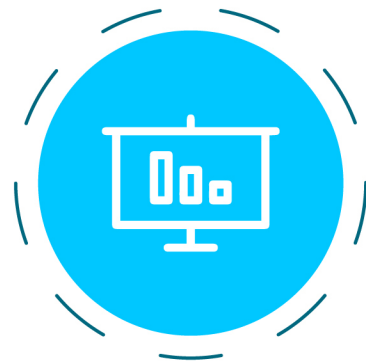
What is the significance of having a portfolio? Even with your skills, resume, degree, and experience, a portfolio provides numerous benefits, such as:



An unmatched first impression on clients & employers.



More holistic than a resume.



A professional and credible way to showcase your work.



Imperative to build your online presence.



Easily shared with a link.



A collection of various types of media files (audio, video, etc.).

Job Trends & Projections for Data Analysts

The rising demand for big data professionals is an undeniable fact. As we chart through the age of big data, the requirement for a workforce with data skills is also increasing.

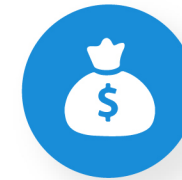
If you are thinking about building a career as a data analyst, you should know about the following trends in the data analytics job market.



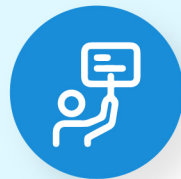
The global big data market is expected to rise to USD 103 billion by 2027.



According to the US Bureau of Labour Statistics, there will be 11.1 million jobs in data analytics by 2026.



A wide range of jobs, such as finance analysts, marketing analysts, business intelligence analysts, operations analysts, etc., are emerging in big data.



The demand for data analysts will grow up to 26 percent by 2026.

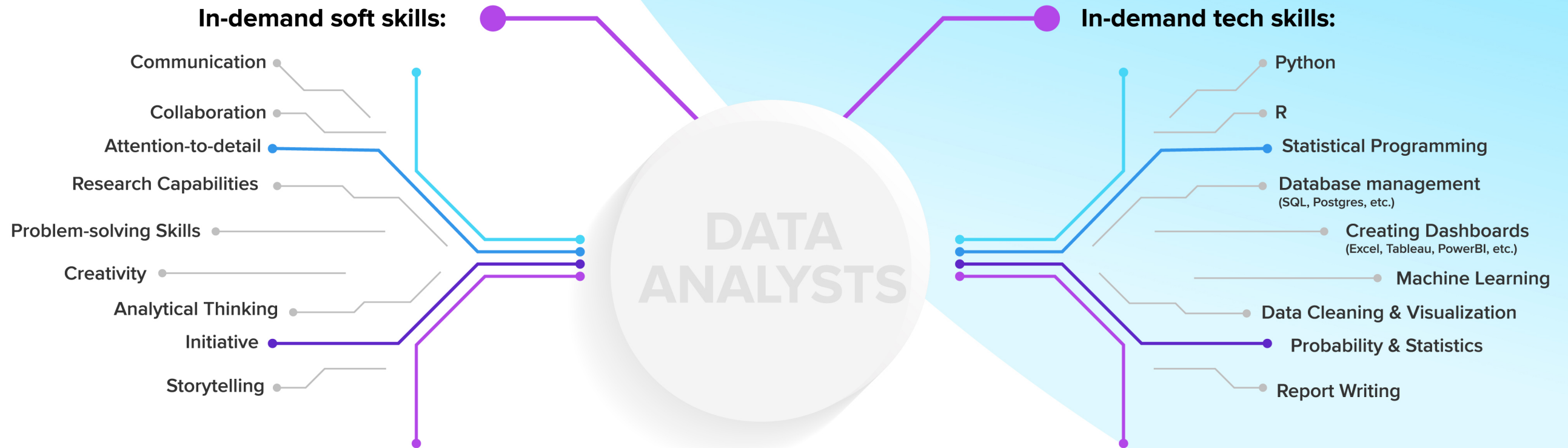


Experience, certifications, projects, & internships are becoming the real game-changers.



Storytelling and Data visualization are listed as required skills in the majority of data analyst job postings.

Industry-Relevant Skills for Data Analysts





What Should a Data Analytics Portfolio Demonstrate?

The right question to ask is – ***What do the employers seek?***
That is what can help you land a data analyst job.
Your portfolio needs to ***demonstrate relevance.***

A relevant educational background shows that you had a vision from the beginning to be in the job you are applying for. A bachelor's in Data Science, Computer Science, or Finance is most relevant. However, given a wide-ranging application of data skills, other degrees can be just as relevant.

If you don't have a formal data science education, you may feel overwhelmed about pursuing a career in this field. However, there's no need to worry. You can pursue professional certifications that will equip you with the necessary skills and expertise to excel as a data analyst. These certifications will provide you with the competencies needed to succeed in a data analytics role and enhance your professional portfolio with relevant and valuable credentials.

Your portfolio must demonstrate the two most underrated soft skills – **initiative & learnability.**

If you are taking on projects & internships without any credits requirement from your curriculum, it shows not only your interest but also that you take initiative on your own. As you take on more diverse projects, your skillset also broadens, showcasing your learnability.

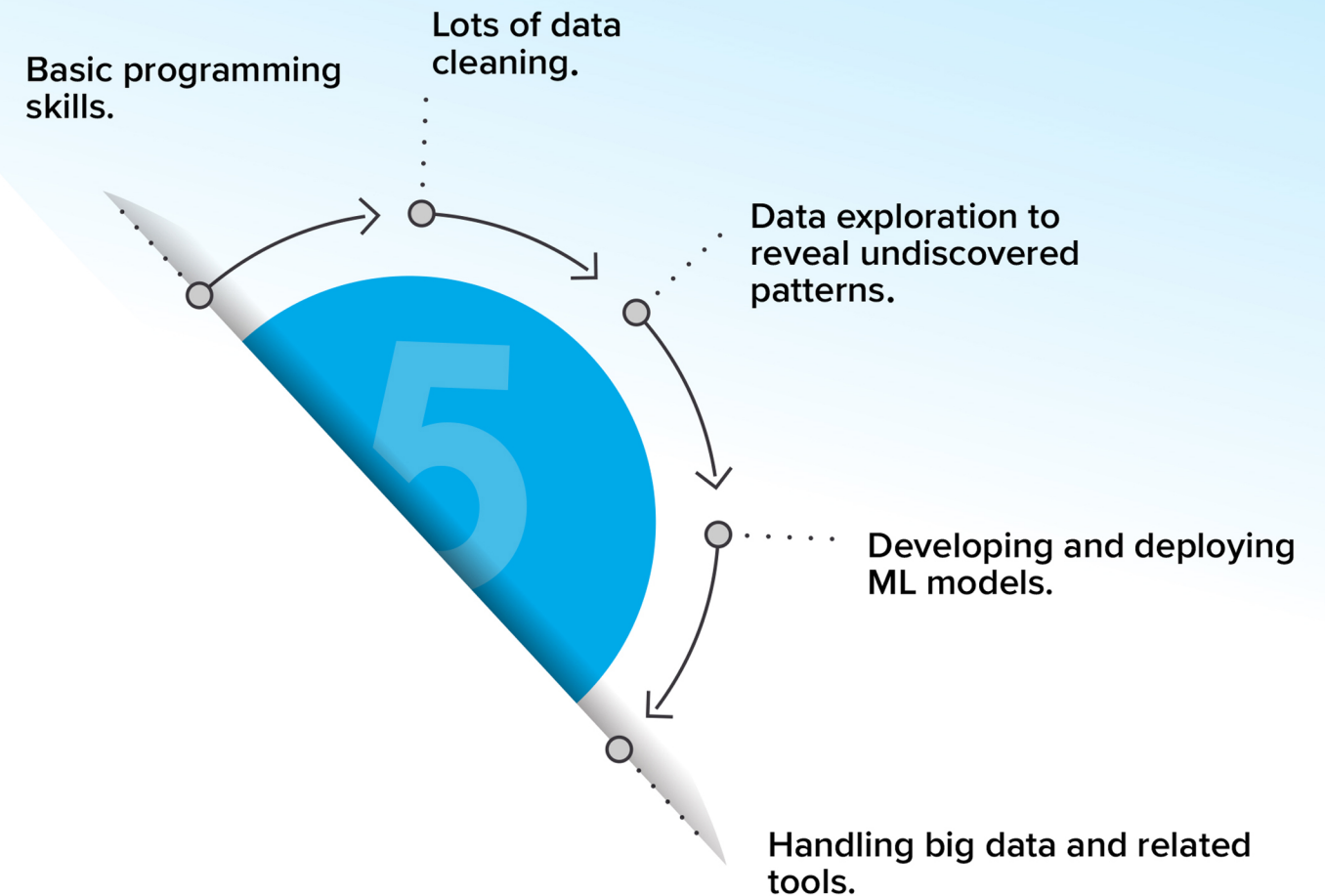
To be a successful data analyst, you should demonstrate proficiency in the following skills – **Data Cleaning, Exploratory Data Analysis (EDA), Data Visualization, Machine Learning Basics, and Statistics.**

It should also feature projects and relevant experiences that demonstrate your proficiency in these areas. These projects are crucial as they enable you to harness the power of data to drive meaningful insights and informed decision-making. Let's explore data analytics projects in greater depth, including their various types.

What are Data Analytics Projects?

A project in data analytics involves a series of tasks that must be completed to achieve the desired outcome. It uses current and past project information to facilitate effective decision-making for project delivery. This involves collecting, organizing, cleaning, managing, processing, and visualizing data to address real-world issues.

A data analytics project must include:



Importance of Doing Data Analytics Projects

Doing projects on your own is a challenging but rewarding feat. That is why companies are closely looking at candidate's profiles for note-worthy projects they have done before applying for a position.

Here's how engaging in such projects can be a stepping stone for your career in data analytics:

They enhance your analytical abilities by requiring you to analyze and make sense of complex data sets. This improves your trend, pattern, and insight identification skills, allowing you to mitigate risk and handle setbacks more effectively.

Working on self-driven projects demonstrates interest and initiative in a candidate's profile, indicating that you have a genuine passion for data analysis and a willingness to learn and grow in the field.

Data analytics projects provide a better way for hiring managers to assess your skill level before offering you a chance to present yourself, giving them an idea of your potential as a data analyst.

Projects offer an opportunity to showcase your experience in using various tools like SQL, Excel, Python, R, SAS, etc. These are skills that companies look for when posting a data analyst job, making you a valuable candidate for such roles.

Engaging in data science projects also improves your communication skills, allowing you to effectively communicate data insights to stakeholders who are unfamiliar with data analysis. This enhances your ability to articulate the value of data-driven decision-making, which is a critical skill for any data analyst.

The background of the slide features a hand interacting with a futuristic digital interface. The interface displays various data visualizations, including a bar chart, a line graph, and a pie chart. The overall aesthetic is high-tech and data-driven, with a blue color palette and glowing elements. The hand is positioned in the center-right, pointing towards the data visualizations.

Recommended Data Analytics Projects for Beginners

So, you just got into data analytics, have learned the basics of the field, and now want to practice your hand at it. Here's a list of the projects you could build as a beginner in data analysis:

1. Scraping Data From Job Portals & Subreddits

While learning data analysis through guided projects, you get freely available datasets to work with. But it is not so in the real world. There are people whose job is to source and scrap this data from somewhere. In some organizations, it is the job of a data analyst to collect such data.

Hence, data collection becomes an essential skill for you to learn. An appropriate way to learn data collection as a beginner is to try your hands on a scraping project. There is no dearth of resources on the internet that will guide you through this project.

Problem Statement: Learn data collection by scraping data from job portals and subreddits.

A general guideline to complete this project looks like this :

- a. Select a job portal or a subreddit you want to scrap.
- b. Select the toolkit you want to work on - Python, Scrapy, & Selenium can be good choices, but it's better to do your research.
- c. Write the script to scrape the data.
- d. Make sure to collect the data in an organized fashion.

In a real-life job, you might be doing this a lot if your primary function in the organization is to collect data. Still, data collection is a skill that every data analytics professional should know.

2. Exploratory Data Analysis on a Dataset

Exploratory data analysis is not one single project but a class of projects. In EDA, you explore a dataset, calculate statistical parameters, visualize it, and discover unknown patterns and information. While EDA projects don't seem like much work, they are undeniably helpful to learn real-life data skills.

Problem Statement: Understand the underlying structure of the data and generate initial ideas for modeling or further analysis.

Here's how to go about doing an EDA project:

- a. Select a dataset to work with.
- b. Start exploring the data for anything you think could be valuable/meaningful in any scenario. Calculate its statistical parameters.
- c. Start visualizing various features to bring out undiscovered patterns.
- d. Write down a report of what you discovered, drawing inferences from it.

3. Housing Prices Prediction

Housing price prediction is an introductory project to familiarize yourself with linear regression. It is generally the first prediction model that machine learning practitioners and data analysts make while beginning to learn AI.

The problem goes like this – There is a housing dataset containing a list of house features and their prices. Given this data, you must make a model that can predict the cost of other houses. With the boom in the real estate sector, this problem became a popular one for data guys to work on.

Problem Statement: Building a prediction model to assist home buyers and sellers in determining whether they are receiving a fair deal, without the need for a brokerage firm.

Follow these steps to build yourself a housing price prediction model:

- a. Read the theory behind linear regression.
- b. Select the toolkit you want to work with. Python with Scikit learn is a popular choice.
- c. Grab the housing price dataset. While this is generally inbuilt into ML libraries, you can find it all over the internet with a simple Google search.
- d. Build, Train, and Test the prediction model

4. Forest Fire Prediction

Managing forest fires is difficult due to their unpredictability and late discovery, resulting in inefficient infrastructure development and resource allocation.

To predict forest fires accurately and facilitate a timely response, a model can be developed using machine learning algorithms such as K-means clustering. Incorporating precise meteorological data that cause forest fires can enhance the predictions, making them more targeted for a better response.

Problem Statement: To accurately predict forest fires to facilitate a timely response.

These steps should help you in getting started with your Forest Fire project:

- a. Read the theoretical approach behind K-means clustering.
- b. Choose a toolkit.
- c. Find the forest fire dataset.
- d. Build, Train, and Test the prediction model.

5. Exploring Yelp Data for Good Restaurants

This one is one of those projects that can help people out in their day-to-day lives. So, making this project showcases your skills while solving a real-life problem. You can either use this project to discover good restaurants for yourself or make a recommendation system for others to use. However, building a recommendation system can be a little more challenging for beginners.

Problem Statement: To develop a restaurant recommendation system that uses data analysis to help users discover high-quality dining experiences.

Here's what you should do:

- Revise the theory behind building recommendation systems.
- Choose the appropriate dataset.
- Choose a toolkit as per your preference.
- Build, Train, and Test the model.

Recommended Data Analysis Projects for Mid-Level Professionals

Advanced-level data analytics projects often leverage machine learning to a great extent. If you are done with the basic projects, you already have enough skills to be a data analyst. However, having strong AI skills can help you stand apart from the competition. These projects will also serve to bring you more confidence in terms of handling data in different formats & making decisions with AI.

1. Chatbots

Chatbots are already at the peak of their popularity as they allow businesses to serve customers without any massive resource deployment. You won't be building chatbots for your organization as a data analyst. Still, this is a notable project as it allows you to analyze text data through NLP. Additionally, chatbots are becoming increasingly helpful in organized data collection.

Hence, with an advanced level of understanding of data analytics tools and concepts, a chatbot is something you should try your hands on. If you make a narrowly focused chatbot, like a ticket registration bot, it is called a domain-specific bot. If you make a broadly focused chatbot, it can converse with you about anything. However, this kind of chatbot is harder to make and needs a lot of training data.

Problem Statement: Making business-customer communication easy and productive.

Here's what you should do:

- a. Choose whether you want to make a domain-specific or open-domain chatbot.
- b. Revise the theory behind RNNs & NLP.
- c. Choose a toolkit as per your preference.
- d. Grab a dataset of your choice.
- e. Build, Train, and Test the model.

2. Sentiment Analysis

Sentiment analysis is a crucial real-life application of data analysis and machine learning that companies cannot do without. It involves the application of NLP, text analytics, and concepts of computational linguistics.

Companies need to know how their customers feel about them, their products, and their vision. Users always leave a lot of their opinions on social media platforms, forums, review sites, etc. This project entails scrapping that data, cleansing, organizing, and analyzing it to discover the sentiments of the customers.

It helps in reputation management, product development, and brand positioning. The sentiments are discovered by running text analytics on the reviews collected.

Problem Statement: The goal is to cleanse, organize, and evaluate data to uncover customer sentiments.

The following steps should help you get started:

- a. Revise your NLP concepts.
- b. Learn the basics of computational linguistics.
- c. Choose a toolkit you prefer (R or Python).
- d. Choose a data set of your choice as you can analyze Amazon reviews, Paper Reviews Dataset, Facebook comments, Twitter comments, etc., the choice of the dataset depends on you.
- e. Build, Train, and Test the model.

3. Recommendation Systems

Are you noticing lately that everything on the Internet is what you want to see? Amazon somehow magically pulls out the products you like, or Netflix keeps you recommending shows that you cannot stop binge-watching?

It's recommendation systems in action. A recommendation system is an AI model that helps you suggest the best possible product/service to the customer. Recommendation systems generally use clustering algorithms to form a cluster of customers determined by useful features that drive user behavior.

Problem Statement: To build a recommendation system that suggest the best products/services to a customer to increase sales.

To get you started:

- a. Decide on the recommendation system you want to develop. For example, it can be a product recommendation system like the one used by Amazon or a movie recommendation system used by Netflix.
- b. Find the dataset for your project.
- c. Learn more about recommendation systems and clustering algorithms.
- d. Build, Train, and Test your model.

4. Speech Recognition Through Emotions

Convenience and Personalization are two critical factors to a brand's success in today's time. It makes moving toward speech-based touchpoints seem like the next logical step. However, recognizing the emotions behind a speech remains a challenge for machines without human intervention.

When we converse, we don't only communicate an idea but also our emotions. These emotions put the statement in the proper context. Sarcasm sounds different than an affirmation - something machines struggle to identify without an intelligent speech recognition system. This project aims to identify emotions from several audio files in a dataset.

Problem statement: Recognize the exact emotions & context behind statements to serve user needs efficiently.

Follow along to get started:

- a. Choose a Toolkit - Python, Librosa, Numpy, Scikit-Learn, and PyAudio packages can help you out.
- b. Find a Speech Emotion Recognition dataset.
- c. Build, Train, and Test your model.

5. Credit Card Frauds

Credit card frauds are possibly the most common type of financial fraud. It was hard to handle these frauds with limited technology in the past. But the emergence of data analytics has presented us with a chance to fight back.

The AI algorithm analyses the usual patterns in customers' spending behavior. Whenever there is an outlying behavior, that transaction gets denied, and the user receives a prompt to go through verification. Thus, any fraudulent behavior stands stopped in its tracks.

Problem Statement: Detect fraudulent payment activities with the help of AI.

Here's how you can get started with this project:

- a. Revise your classification & clustering concepts.
- b. Choose a toolkit
- c. Source the dataset from the Internet.
- d. Build, Train, and Test your model.



Other projects that you can try your hands on

Fake News Detection

Uber Data Analysis

Traffic Signs Recognition

Loan Prediction

Detecting Parkinson's Disease

Crime Analysis

Job Recommendation System

Customer Segmentation

Traffic Prediction

Stock Market Prediction

Find freely available datasets for you to work on

Kaggle

Google Dataset Search

UCI Machine Learning Repository

Global Health Observatory Data Repository

Data.gov

CERN Open Data Portal

Earth Data

Datahub.io

Must-have Data Analysis Tools in Your Arsenal

Apart from knowing the analytics concepts, data analysts need the right combination of tools to function successfully. The following list of tools won't only help beginners in their data quest but is also a helpful resource for professional analysts who want to increase their work efficiency.

1. Python

Python has skyrocketed in popularity in the last five years. Its compatibility with data analytics and a rich library to handle data has been a major driving force behind its surging demand. It's a general-purpose language making it a suitable choice for a wide range of big data tasks.

Python's libraries like Pandas and Numpy facilitate efficient data handling, cleaning, and organization tasks. Similarly, other modules like Matplotlib and Plotly are suitable for data visualization.

The next logical step in this direction is machine learning, and Python supports the most popular ML/AI libraries like Scikit-Learn, TensorFlow, and Keras.

Additionally, Python is simple to learn and use, with an easy learning curve. So, anyone looking to get started with data science can do so without much problem.

2. R

R is an open-source programming language used for statistical programming and data mining. As R has been designed especially for handling data, it forms a perfect tool for data analytics. After Python, it is the most popular choice among data-driven organizations and professionals alike.

It has a graphical user interface making it an easy-to-use tool. It didn't start this way, but it has now evolved to become more user-friendly, given that people from all backgrounds are using it.

As a data analyst, you can use R for statistical testing, text mining, forecasting, data mining, what-if analysis, time series analysis, sentiment analysis, etc. Data professionals around the globe use it for building state-of-the-art models to solve real-life problems. If you are switching to a career in data analysis, it's for the best that you have R in your arsenal.

3. SQL

When it comes to databases, SQL is your go-to language. It can seamlessly interact with multiple databases at once, making querying data a piece of cake

It has an easy learning curve making it possible for everyone to learn to write even complex instructions with ease. It is powerful; which shows in the intuitive data analytics dashboards that one can build and use for interacting with databases.

What's even better is that you can use it with the popular analytics languages like Python, Scala, and Hadoop. From data warehousing to querying to running analytics, SQL holds its head high as a must-know tool for data analysts.

4. Jupyter Notebook

Jupyter Notebook is an open-source, free-to-use web tool that combines inputs (code) and output in a single place. It makes a useful tool in a data analyst's toolkit due to its ability to run codes, visualize data, solve mathematical equations, do computations & statistical modeling, etc.

Currently, it supports over 40 programming languages but is highly popular among python practitioners.

While the Jupyter Classic Notebook makes a good data analysis tool, JupyterLab has been designed with advanced capabilities to manage workflows like AI modeling, data science, and scientific computing.

5. Tableau

Tableau is a data analytics tool designed to help businesses better harness their data. It allows you to prepare data through cleaning and organizing while also offering analysis and collaborative capabilities. Its actual strength lies in its easy-to-use data-rich visualizations, which make it easier for everyone to draw inferences, ask questions and share insights across the organization.

Business organizations are using Tableau extensively in their data analytics operations. Thus, if you are looking to land a job as a data analyst or business intelligence analyst, Tableau becomes an indispensable tool for you.

6. Excel

Excel is one of the legacy tools used extensively for data cleansing, organization, and analysis purposes. Over time, the tool has also evolved to keep pace with the changing requirements of individuals and enterprises.

The Analyze Data feature makes data analysis very easy and intuitive. Data analysts find this feature useful as they can analyze data through natural language without writing complicated queries or mathematical formulas.

This advantage makes Excel an excellent choice for enterprise users. That's why Excel ranks high in data analyst job profile requirements. If you are looking to become a data analyst, it is good to have some experience with Excel.



Additional Resources That Data Analysts Can Use

1. GitHub

In present times, collaboration and community contribution rank high on employers' checklists. To that end, GitHub is a popular choice among data analysts to showcase their work.

You can add your projects to repositories and share them among other data science practitioners for collaboration and peer reviews. If you have multiple repositories, it is better to organize them on your website. With GitHub Pages, you can make a simple website that showcases your work.

Companies usually expect you to have a GitHub profile demonstrating your work if you are in a tech profile.

2. Kaggle

There's hardly any data science professional who hasn't heard of Kaggle before. Kaggle is a machine learning competition website that hosts a wide range of datasets for you to build projects. It also allows you to source data from an external cloud database like Google Cloud.

The data problems you need to solve on Kaggle are complex, interesting, and close to what you would be solving in real-life data analytics roles. Additionally, there are leaderboards to compare your performance against other users.

If you want to make an impressive portfolio for your potential employers, Kaggle is something you should not miss out on.

3. Google Cloud

Most businesses today are shifting to cloud data storage. And these are the businesses that you want to thrive in as a data analyst. That's why it is good practice to get some hands-on experience with Google Cloud beforehand.

As Google Cloud offers a free usage tier, you can practice and build models on it for free. To build a project on Google Cloud, you will need to learn BigQuery. Once you are through the basics to access the publicly available data on BigQuery, you can use Google Cloud's AI hub for building projects.

4. DAGsHub

DAGsHub is a collaborative platform for managing and sharing data analysis projects. It provides version control, project management, and collaboration tools specifically designed for data science workflows. With DagsHub, data analysts can easily track changes to their code, data, and models, collaborate with other team members and showcase their work to the broader data science community.

DagsHub also supports reproducibility and transparency in data science by making it easy to share code, data, and results, and by providing tools for tracking experiments and reproducing results. It is built on top of Git and other open-source tools and is free and open to anyone.

How to Land a Job in the Field of Data Analytics?

01

Relevant degree

The most efficient way to become a data analyst is by majoring in a relevant subject such as Computer Science, Data Science, Economics, Statistics, or Finance. However, since data analytics has various offshoots like operation analytics, market analytics, etc., other related degrees can also be relevant. The primary skill required is the ability to handle data effectively.

02

Professional Certifications

Skills are considered to be the best currency when you walk into any job market. Whether you are a fresher or a mid-career professional, upskilling via certifications and courses is always an excellent way to succeed and gain a competitive advantage over others. According to Payscale, obtaining a certification alone results in a 75 percent average salary rise for professionals. The likelihood of being chosen by hiring managers during the screening process will significantly improve after completing a professional data analytics certification program.

03

Be technologically sound

No one is analyzing data on papers anymore. You need to be proficient in the tools of the trade like Python, R, Tableau, GitHub, Jupiter Notebook, Excel, etc. Even if you have not majored in CS, being tech-savvy is essential.