**Introduction to Exploratory Data Analysis (EDA)**

Today, we're diving into the fascinating world of Exploratory Data Analysis, or EDA for short. EDA is a crucial step in the data analysis process that helps us uncover insights, patterns, and anomalies hidden within our data before we start building complex models or making critical decisions. It's like the detective work of data science, where we explore and understand our data to make informed choices. Let's get started!

**Purpose of EDA**

Why do we need EDA, you ask? Well, it's like preparing a canvas before creating a masterpiece. EDA allows us to get familiar with our data, understand its structure, and identify potential problems such as missing values, outliers, or inconsistencies. By doing this, we can make better decisions about preprocessing, model selection, and even feature engineering.

**Techniques in EDA**

EDA offers a toolbox of techniques. Descriptive statistics provide summary information about the data, while visualizations like histograms and scatter plots reveal distribution and relationships. These techniques enable engineers to quickly grasp the essential characteristics of the data.

**1)Data Cleaning**

Before applying engineering solutions, we need clean data. EDA helps us identify and handle missing values, duplicates, and outliers. Just as you'd calibrate your instruments before conducting an experiment, EDA ensures the accuracy of your data analysis.

**2)Visualization in EDA**

Visualization is a powerful EDA tool. It's like creating blueprints before building a structure. Graphs and plots translate complex data into intuitive visuals, helping engineers spot trends and irregularities at a glance.

**3)Identifying Patterns**

In engineering, you search for patterns to solve problems. EDA helps you identify patterns in your data – perhaps cyclic behaviors in time-series data or clusters in multidimensional data. These patterns guide your engineering decisions.

## Prerequisites

The following prerequisites are required to follow the tutorial:

* [IBM Cloud account](https://cloud.ibm.com/registration?cm_sp=ibmdev-_-developer-tutorials-_-cloudreg)
* [IBM Watson® Studio](https://www.ibm.com/cloud/watson-studio)
* [IBM Watson Machine Learning Service](https://www.ibm.com/cloud/machine-learning)

## Steps

1. [Create your IBM Cloud Account and access the IBM Cloud Pak for Data as a Service.](https://developer.ibm.com/tutorials/build-a-logistic-regression-neural-network-using-tensorflow/?mhsrc=ibmsearch_a&mhq=tutorials#create-ibm-cloud-account)
2. [Create a new project.](https://developer.ibm.com/tutorials/build-a-logistic-regression-neural-network-using-tensorflow/?mhsrc=ibmsearch_a&mhq=tutorials#create-a-new-project)
3. [Associate the Watson Machine Learning Service with the project.](https://developer.ibm.com/tutorials/build-a-logistic-regression-neural-network-using-tensorflow/?mhsrc=ibmsearch_a&mhq=tutorials#Associate-the-watson-machine-learning-service-with-the-project)
4. Create an Environment and [Add a Notebook to your project.](https://developer.ibm.com/tutorials/build-a-logistic-regression-neural-network-using-tensorflow/?mhsrc=ibmsearch_a&mhq=tutorials#add-notebook-to-project)
5. [Run the Notebook.](https://developer.ibm.com/tutorials/build-a-logistic-regression-neural-network-using-tensorflow/?mhsrc=ibmsearch_a&mhq=tutorials#run-the-notebook)

### Step 1. Create IBM Cloud account

1. Sign in to your [IBM Cloud account](https://cloud.ibm.com/registration?cm_sp=ibmdev-_-developer-tutorials-_-cloudreg).
2. Search for Watson Studio.
3. Create the service by selecting a region and pricing plan and terms of agreement

4.Click on Launch in IBM Cloud Park as Data

### Step 2. Create a new project

1. **Get started** the Watson Studio service.
2. Click on **a New Project**, and then click on **NEXT**
3. **Name** your project, add a **storage** service and then click on **create**
4. After Click on **Create**. After your project is created, you are directed to a project dashboard, and click on the **Manage tab** it will display the below

### Step 3. Associate the Watson Machine Learning Service with the project

1.Scroll down to click on  **Services&Integrations**  then click on **Associate Services**

2. After clicking on **Associate services** and then appears new tap click on Waston Machine learning tick mark and then click on **Associate**

3.After completing of above things click on **New Services**

4.After clicking new services then appear **new dashboard of services** then click on **AI/Machine learning** and select the **Waston Machine learning**

J

1. After clicking on **Waston Machine Learning** next appear the new dashboard for then **click on Create** of Waston Machine learning

### Step 4. Create an Environment and Add Notebook to project

1. After that click the **“Environment tab”** then click on **“New template”**
2. After clicking on **NEW Template,** it appears **“New Environment box “** create the environment name then **click** **Create**

3.After the creation of a new environment, we click the right side 3dots of the environment, the appear small box consists of new notebook, promote to space, edit , delete out of this we select new notebook

4. After that appears **New Notebook Dialogue box**, In that we select **Notebook from the URL** and next we **create the notebook name** and we paste the below link in the URL box and click create  [https://github.com/Sardaruzma/EDA/blob/main/EDAUpdated.ipynb](https://github.com/IBM/dl-learning-path-assets/tree/main/fundamentals-of-deeplearning/notebooks/Logistic_Regression_with_TensorFlow.ipynb)

### Step 5. Run the Notebook

1.After the notebook is loaded, click **Cell**, then select **Run All** to run the Notebook.

**GitHub link for all the files**

https://github.com/Sardaruzma/EDA/tree/main

**Conclusion:**

In conclusion, Exploratory Data Analysis is a cornerstone of data-driven engineering. It's your initial step to untangle the data's story, enabling you to make informed decisions, build robust solutions, and drive innovation in your engineering projects. So, embrace EDA as your engineering compass, guiding you through the data landscape. Happy exploring!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Day** | **Temperature** | **Ice Cream Sales** | **Age** | **Happiness Score** |
| **1** | **30** | **50** | **25** | **8.5** |
| **2** | **28** | **45** | **32** | **8.0** |
| **3** | **33** | **60** | **28** | **9.2** |
| **4** | **22** | **30** | **9** | **9** |
| **5** | **25** | **40** | **22** | **7.2** |
| **6** | **35** | **70** | **40** | **9.8** |
| **7** | **32** | **80** | **19** | **8.5** |
| **8** | **29** | **55** | **31** | **8.3** |
| **9** | **24** | **35** | **27** | **7.0** |
| **10** | **38** | **800** | **555** | **9.5** |