Healthcare - Predictive Analytics for Patient Care

DATA SCIENCE PROFESSIONAL:

1.**Exploratory Data Analysis** : analyse patient data(such as age, treatment history) to identify patterns in patient readmissions.

Example: performing EDA on lungcancer dataset

```python
import pandas as pd

# Load CSV file from Google Drive
df = pd.read_csv('/content/drive/My Drive/DATASET.csv')

# View the first few rows
df.head()
```

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN	LUNG_CANCER
0	M	69	1	2	2	1	1	2	1	2	2	2	2	2	2	YES
1	M	74	2	1	1	1	2	2	2	1	1	1	2	2	2	YES
2	F	59	1	1	1	2	1	2	1	2	1	2	2	1	2	NO
3	M	63	2	2	2	1	1	1	1	1	2	1	1	2	2	NO
4	F	63	1	2	1	1	1	1	1	2	1	2	2	1	1	NO

We Imported the dataset and made a pandas header for it.

Now we are describing the dataset.

```python
# Basic info about the dataset
df.info()

# Descriptive statistics
df.describe()

# Check for missing values
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 309 entries, 0 to 308
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   GENDER                 309 non-null    object
 1   AGE                    309 non-null    int64
 2   SMOKING                309 non-null    int64
 3   YELLOW_FINGERS         309 non-null    int64
 4   ANXIETY                309 non-null    int64
 5   PEER_PRESSURE          309 non-null    int64
 6   CHRONIC DISEASE        309 non-null    int64
 7   FATIGUE                309 non-null    int64
 8   ALLERGY                309 non-null    int64
 9   WHEEZING               309 non-null    int64
 10  ALCOHOL CONSUMING      309 non-null    int64
 11  COUGHING               309 non-null    int64
 12  SHORTNESS OF BREATH    309 non-null    int64
 13  SWALLOWING DIFFICULTY  309 non-null    int64
 14  CHEST PAIN             309 non-null    int64
 15  LUNG_CANCER            309 non-null    object
dtypes: int64(14), object(2)
memory usage: 38.8+ KB
Unique values in GENDER: ['M' 'F']
Unique values in LUNG_CANCER: ['YES' 'NO']
```

Now Plotting various unique columns(Features) to gain insights

```python
# Analyze the distribution of numerical features
import matplotlib.pyplot as plt
import seaborn as sns

for column in df.select_dtypes(include=['number']):
    plt.figure(figsize=(8, 6))
    sns.histplot(df[column], kde=True)
    plt.title(f"Distribution of {column}")
    plt.show()
```

One such insight is Distribution of **smoking** and **age**

Distribution of AGE
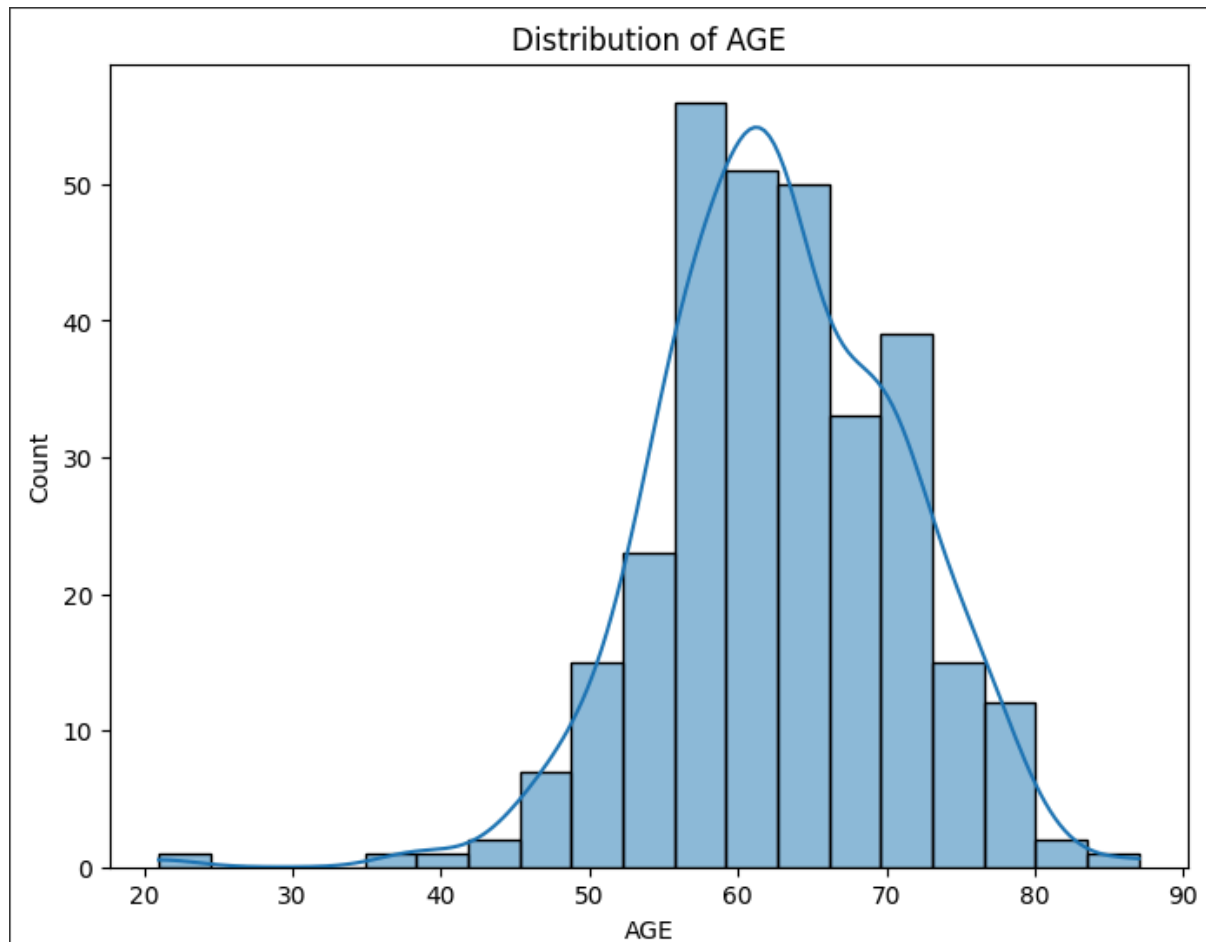
2.:**Feature Engineering**: Create features such as length of stay, previous readmission rates, that can influence patient risk patterns or profiles.

MACHINE LEARNING PROFESSIONAL:

1.**Predictive modelling**: Use various Machine learning algorithms to predict patient readmission .

2.**Identification of High risk patients** using various ML algorithms and patient records.

ARTIFICIAL INTELLIGENCE PROFESSIONAL:

1.**NLP(Natural Language Processing):** Use of NLP in patient notes, doctor notes , observation notes to identify points of interest such as symptoms , disease patterns ,etc.

2.Developing **Deep neural network** models for identification of high risk chronic diseases such as cancer from complex reports such as MRIs.