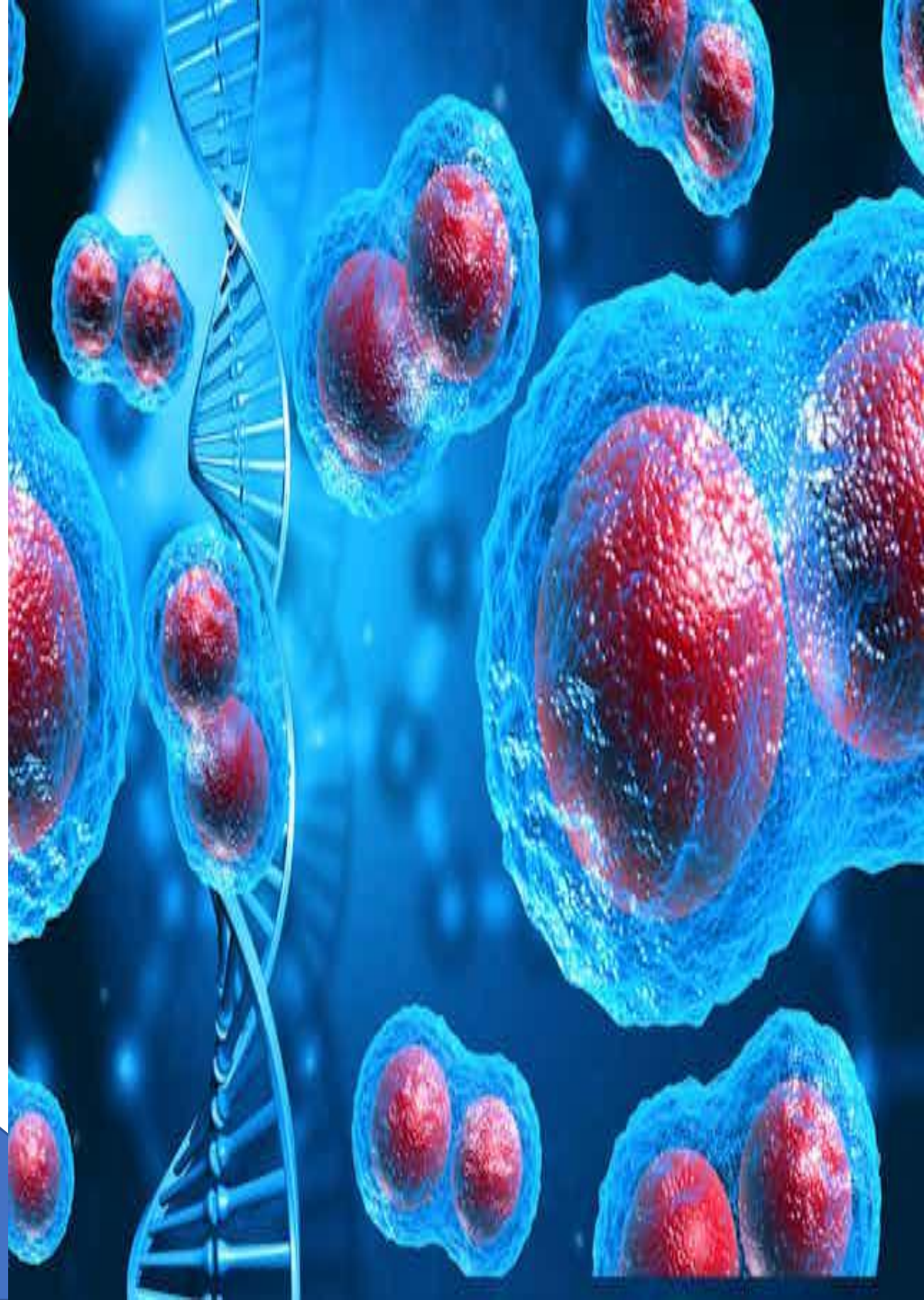
UAE CANCER PREDICTIVE MODELLING PRESENTATION



PRESENTATION BY:

DATA ANALYSTS - GROUP 3:

- **OLUOMA ILOBAH**
- **IYANU ADELE**
- **MARY EDEH**

PROJECT OBJECTIVE

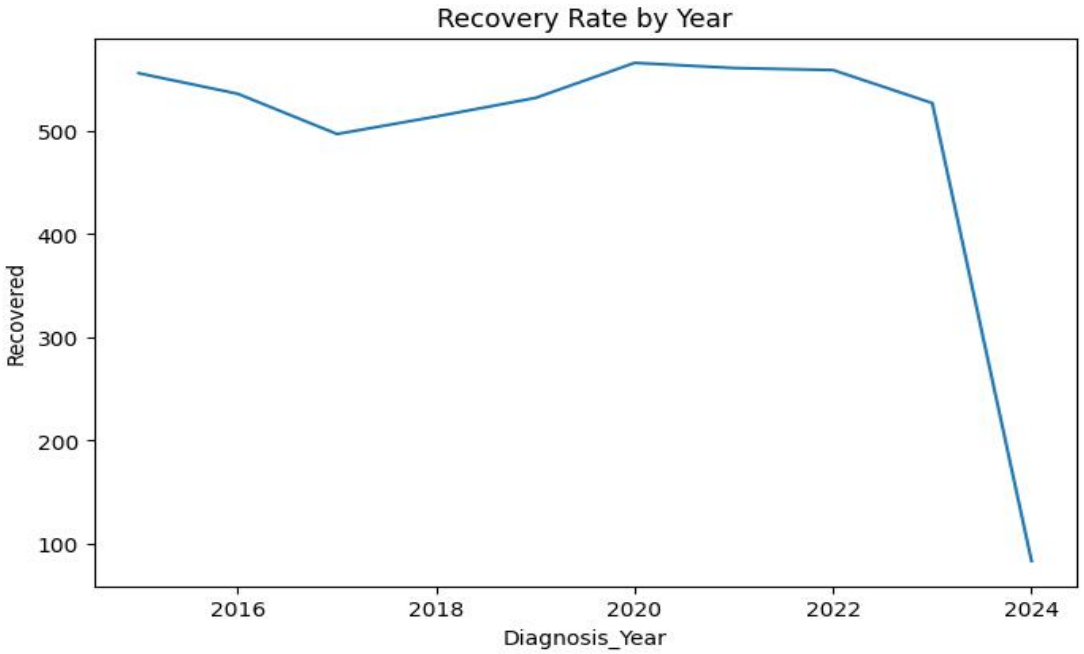
To analyze cancer patients data UAE using machine learning for two main purposes:

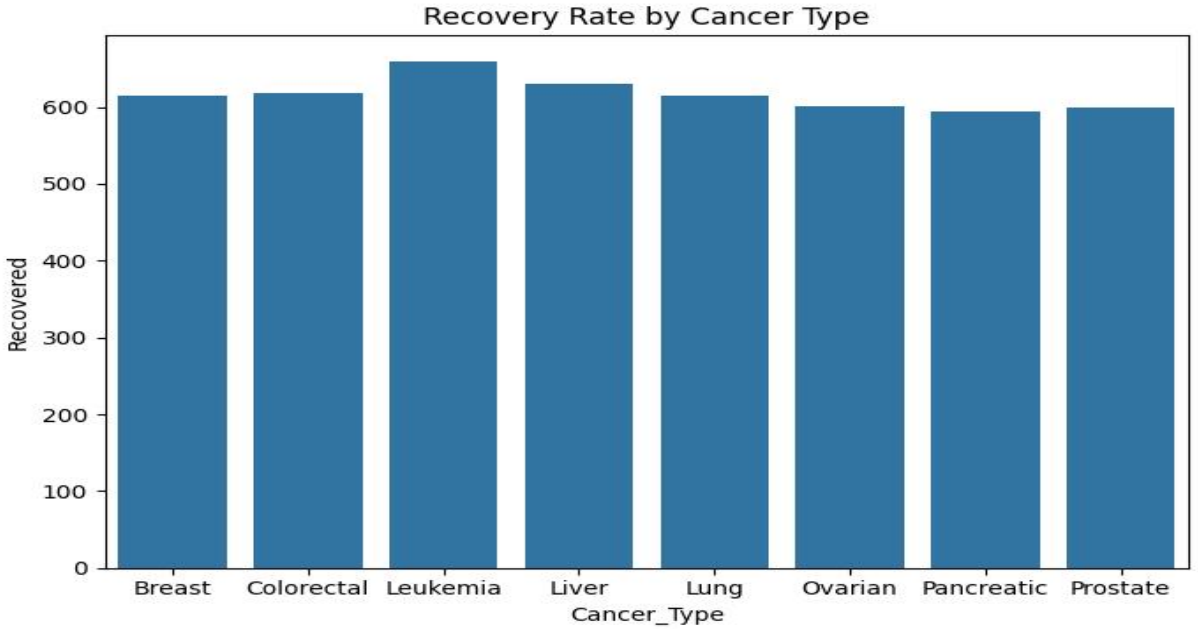
Classification: Build predictive models to identify the cancer stage of patients based on available demographic and medical data

Clustering: Discover hidden patterns and patient subgroups using unsupervised learning to support personalized treatment strategies and healthcare insights

EXPLORATORY DATA ANALYSIS

Key Findings





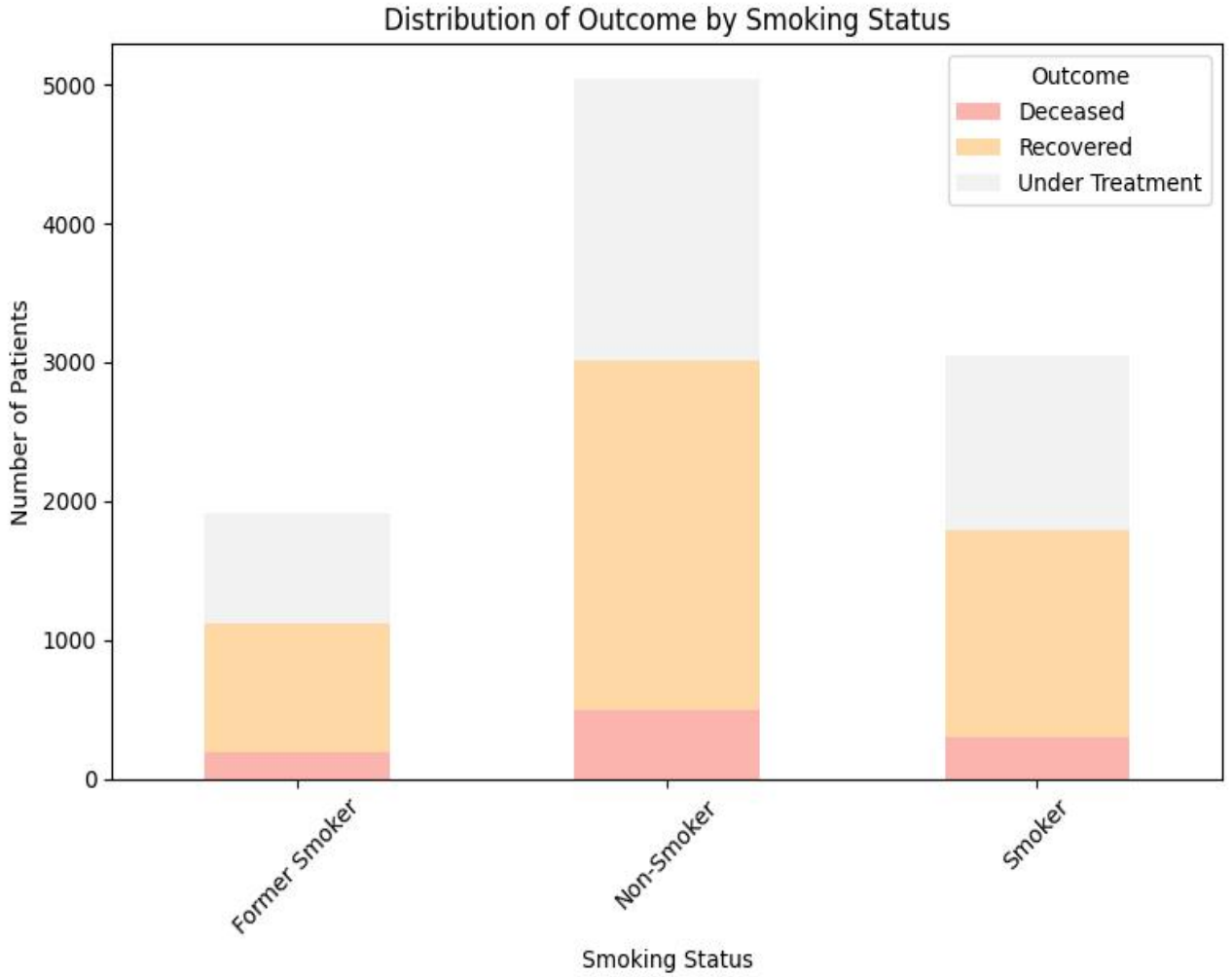
Highest in 2020 (566), 2021 (561), and 2022 (559)

Slight dip in 2023 (527) - which could be due to incomplete records

Highest among Leukemia (660), Liver (630), and Breast (614) patients

EXPLORATORY DATA ANALYSIS

Key Findings

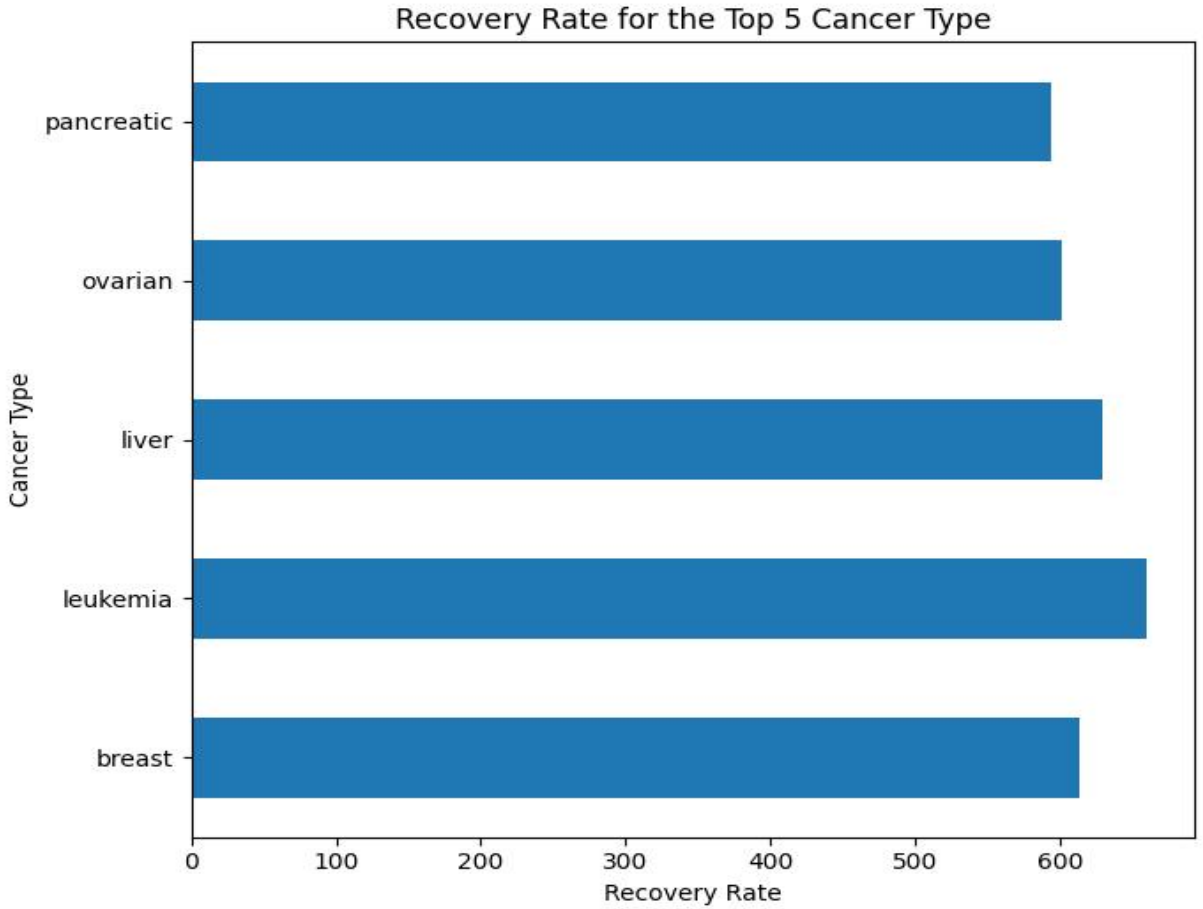


Outcome	Deceased	Recovered	Under Treatment
Former Smoker	195	929	793
Non-Smoker	492	2515	2031
Smoker	305	1487	1253

EXPLORATORY DATA ANALYSIS

Key Findings

Top 5 most common cancer types:
Leukemia (1314), Liver (1263), Ovarian (1259), Pancreatic (1243), Breast (1241)



Cancer Type	Deceased	Recovered	Under Treatment
Breast	120	614	507
Leukemia	127	660	527
Liver	115	630	518
Ovarian	128	601	530
Pancreatic	145	594	504

PREDICTIVE MODELLING

Model to predict Cancer stages I - IV:

FEATURES USED

- Demographics: Age, Gender, Nationality, Ethnicity
- Lifestyle: Smoking Status
- Medical History: Comorbidities, Cancer Type, Weight, Height

MODELS TRAINED

- Random Forest Classifier
- Logistic Regression
- Support Vector Classifier (OneVsRest and OneVsOne)
- XGBoost Classifier

Random Forest (with SMOTE): Accuracy = 30% (highest)
After tuning: 29%

CLUSTER ANALYSIS

We carried out unsupervised learning to uncover patients subgroups:

Elbow Method → Best at k = 2

Evaluation Metrics

- ❑ Silhouette Score: 0.0934 (low cohesion)
- ❑ Davies-Bouldin Index: 2.8964 (high similarity)
- ❑ Calinski-Harabasz Index: 1171.59 (moderate dispersion)

PCA Insights

- ❑ PC1: BMI (0.709), Weight (0.622), Height (0.331)
- ❑ PC2: Height (0.649), Weight (0.356)
- ❑ Age had minimal impact

CLUSTER ANALYSIS

Cluster Summary

Cluster	Age	BMI	Weight	Height	Recovery Rate
0	53.99	29.67	80.85	1.65	50.3%
1	53.15	20.11	60.03	1.73	48.4%

- **Cluster 0: Higher BMI/weight, slightly better recovery**
- **Cluster 1: Leaner profile, slightly lower recovery**

OUR RECOMMENDATIONS

Recommendations
Collect richer clinical/biological data
Apply deeper dimensionality reduction or feature selection
Explore non-tabular approaches