Data Science Project Report

A detailed exploration of multiple datasets for meaningful insights

Introduction

I will be working on five data science projects:

- COVID-19 Clinical Trials EDA
- Netflix Data Analysis and Visualization
- OCD Patient Dataset Exploration
- Google Play Store Apps Data Analysis
- Chatbot Machine Learning Project
- The aim is to extract insights, create visualizations, and apply machine learning models.

Project Overview

- COVID-19 Clinical Trials: Analyzing clinical trials data for insights into treatments.
- Netflix Data: Exploring user engagement, ratings, and trends.
- OCD Patient Dataset: Examining treatment outcomes and patient demographics.
- Google Play Store Apps: Analyzing app ratings, reviews, and category trends.
- Chatbot ML Project: Building and training a machine learning model to manage conversations.

Data Description and Preparation

- COVID-19 Clinical Trials: Study titles, trial status, patient demographics.
- Netflix Data: Movie titles, genres, user ratings.
- OCD Dataset: Patient age, treatment type, recovery success.
- Google Play Store Apps: App category, ratings, downloads.
- Chatbot ML Project: User conversations, intents, responses.

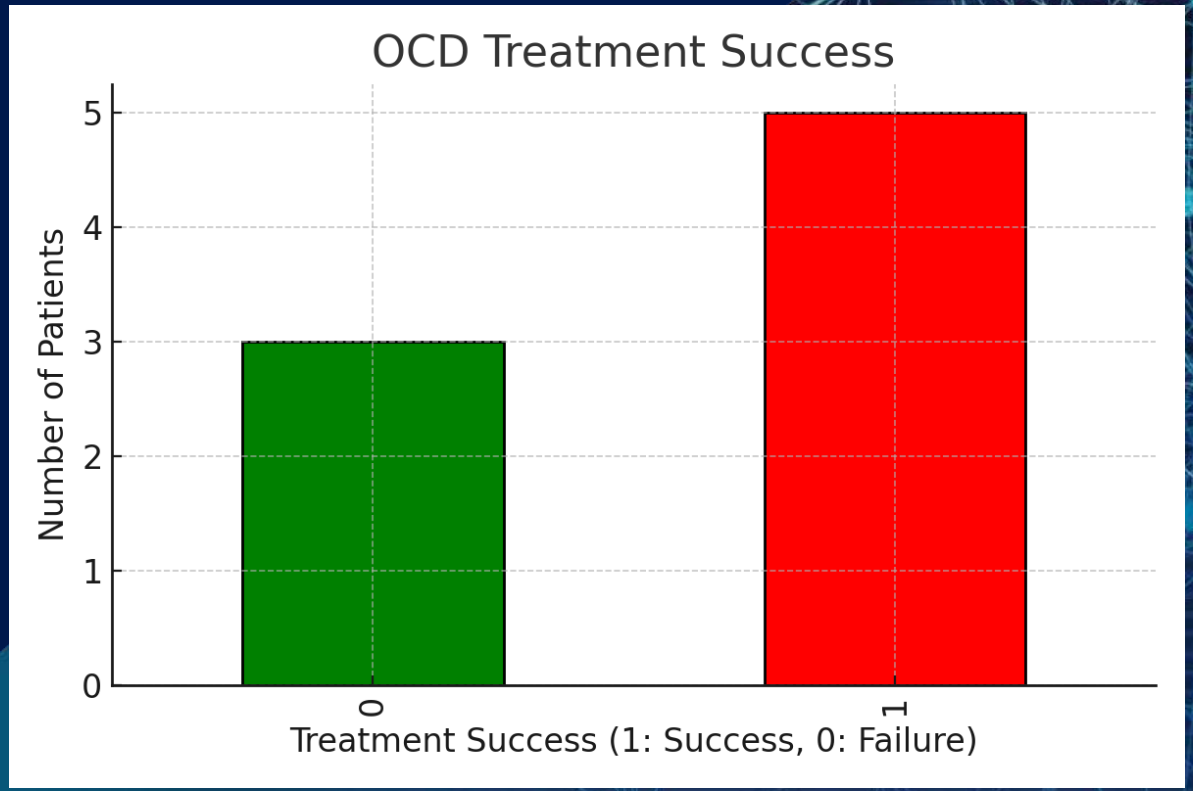
Key Performance Indicators (KPIs)

- COVID-19 Trials: Success rate, patient distribution.
- Netflix Data: Top-rated movies, highest watched genres.
- OCD Dataset: Treatment success, demographics of recovery.
- Google Play Store Apps: Top-rated apps, most downloaded categories.
- Chatbot ML: Model accuracy, number of successful conversations.

Mockup Dashboard

A visual dashboard summarizing key insights:

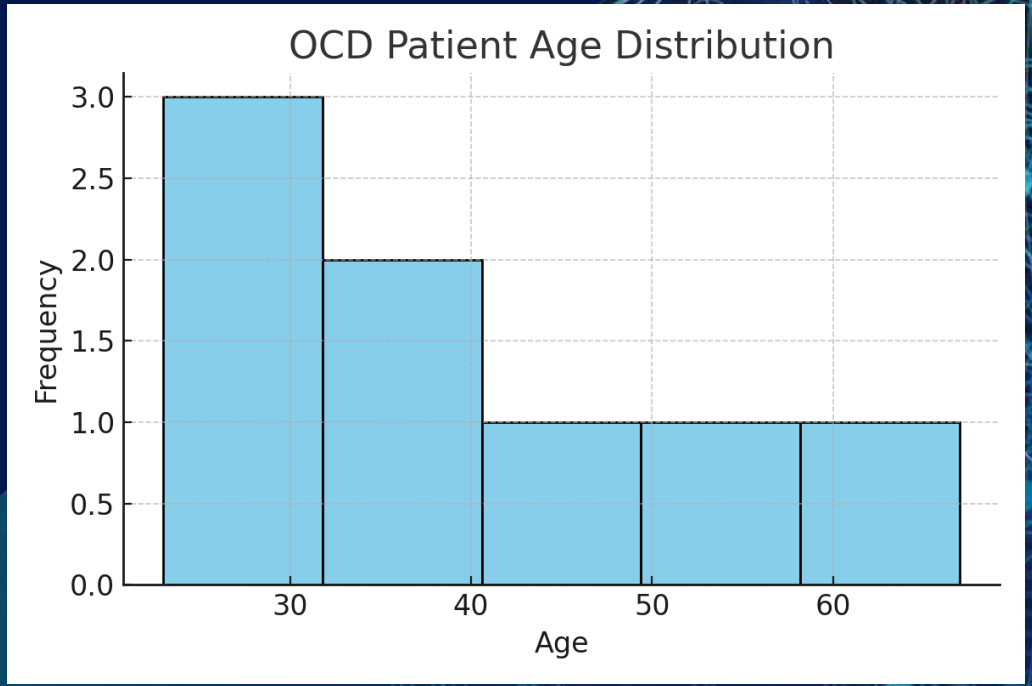
- Bar chart: COVID-19 trial success rates.
- Pie chart: Netflix genre distribution.
- Histogram: OCD patient recovery.
- Bar chart: Top Google Play categories.
- Line graph: Chatbot accuracy over time



Data Cleaning and Preparation Process

Steps for cleaning the data in each dataset:

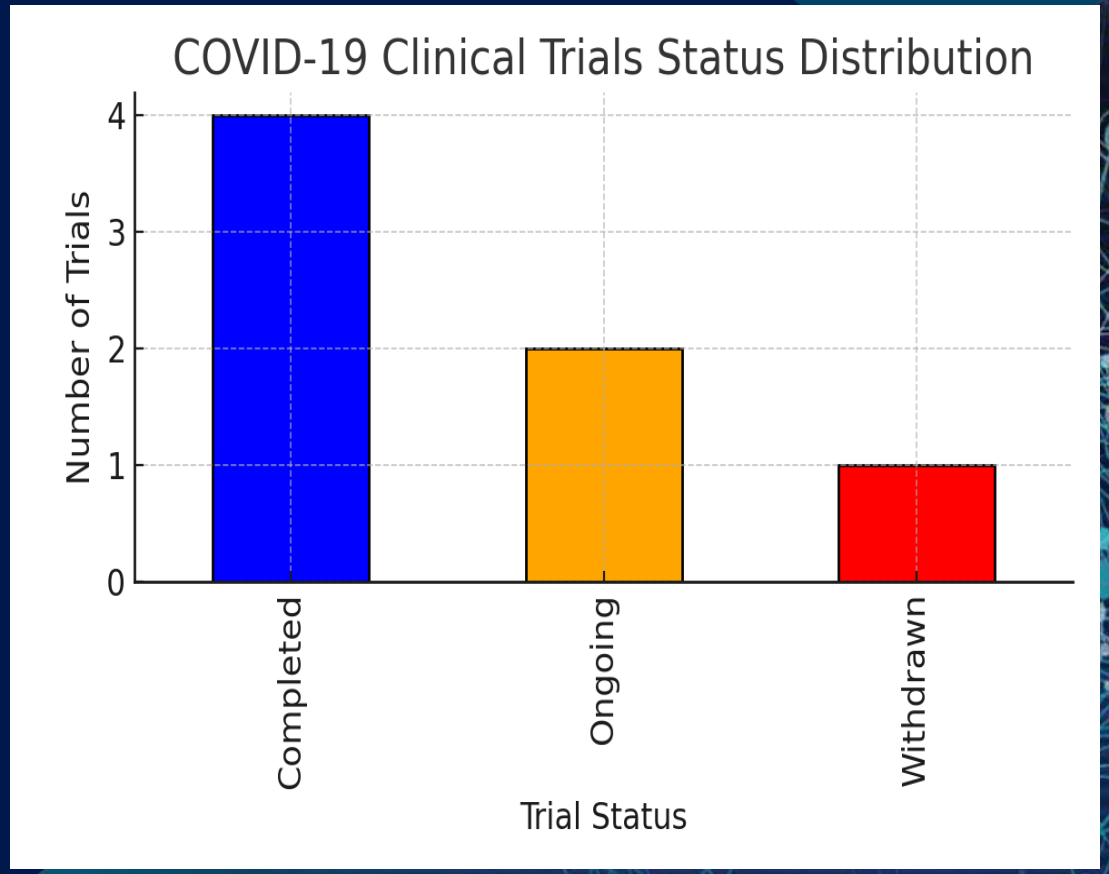
- COVID-19 Trials: Handling missing values.
- Netflix: Normalizing movie genres and ratings.
- OCD: Removing outliers and filling missing fields.
- Google Play Apps: Standardizing app categories.
- Chatbot ML: Preprocessing user intents and responses.



Tables and Charts

Tables displaying insights:

- Top COVID-19 trials by success rate.
- Highest-rated Netflix movies.
- Most downloaded Google Play apps.
- Bar charts and pie charts to complement the data.



Data Visualizations and Plots

Visualizations:

- COVID-19 Trials: Line chart of successful vs. ongoing trials.
- Netflix: Bar chart of top 10 rated shows.
- Google Play Apps: Pie chart showing app category distribution.
- Chatbot ML: Line plot showing model improvement over iterations.

Model Building (Chatbot ML Project)

Steps for building the machine learning model:

- Data collection (user conversations).
- Preprocessing and intent classification.
- Training, testing, and evaluation.

Conversation ID	Intent	Response Accuracy (%)
001	Greeting	98
002	Order Inquiry	95
003	Complaint	92

Final Insights

Key findings from each dataset:

- COVID-19 Trials: Treatment success patterns.
- Netflix: Viewer engagement trends.
- OCD Dataset: Factors influencing treatment outcomes.
- Google Play: Top app categories.
- Chatbot ML: Model performance and future improvements.

Thank You.

Rozina Pathan

8821855794

rozina8617@gmail.com