

PROJECTS

1. COVID VACCINATION DATA ANALYSIS

- **Objective:** Analysing the status of COVID-19 vaccines (Covishield, Cowin,) available, and predicting the completion of all doses of vaccines in India.
- **Steps performed:** data set generation, data cleaning, data preprocessing steps such as handling missing values and removing irrelevant records, visualization, and analyzing trends.
- Utilized **libraries** like **pandas, NumPy, seaborn, matplotlib, and Plotly** to perform data cleaning and explore trends, visualize vaccination progress across countries, and generate interactive choropleth maps.
- Followed **best practices** in analysis, visualization, and accuracy prediction.
- **GitHub:** github.com/saikale/Deck/blob/main/Covid_19_Vaccination_Analysis.ipynb

2. STARTUP FUNDING ANALYSIS

- **Objective:** To analyze the status of startup funding in India, identify top investors, and track industry verticals and cities with the highest investments, thereby predicting funding trends and providing insights into the evolving investment ecosystem.
- **Key steps performed:** Data Collection, Data Cleaning and Preprocessing (Handling Missing Values, Removal of Irrelevant Data, Data Formatting), Visualization & Trend Analysis (Top Investors, City wise Investments, Industry Vertical Analysis, Time-based Trend Analysis), Prediction & Forecasting.
- Utilized **tools** like Pandas, NumPy, Seaborn & Matplotlib.
- Followed **best practices and structured approach** to data cleaning, analysis, and visualization.
- **GitHub:** github.com/saikale/Deck/blob/main/Startup%20Funding%20Analysis.ipynb

3. EXPLORATORY ANALYSIS OF REAL ESTATE IN BANGALORE

- **Objective:** To analyze the Bengaluru (metropolitan city in India) housing market, understand the factors affecting house prices, and develop a predictive model for estimating property values, hence, to assist potential buyers or investors in making informed decisions.

- **Key steps performed:** Data Collection, Data Cleaning and Preprocessing (Removing irrelevant columns, handling missing values, data extraction, data conversion, one-hot encoding), Visualization & Trend Analysis (histograms, scatterplot).
- **Modelling** (Train-test split, Linear Regression Model), Model Evaluation (Evaluated the model's performance using Mean Squared Error (MSE) and R-squared metrics)
- Utilized **tools like libraries:** Pandas, NumPy, Scikit-learn for modeling, and Matplotlib & Seaborn for data visualization.
- **GitHub:** github.com/saikale/Deck/blob/main/Real_estate_Project.ipynb

4. INTERVIEW FEEDBACK SYSTEM USING CHATGPT

- **This project was part of a company hackathon. This was a group work where I performed a key leadership role.**
- **Objective:** To develop an automated system that generates interview feedback based on the interview transcript using the OpenAI GPT model. This system aims to assist recruiters or interviewers in gaining insights into the interview process and provide structured feedback for candidates. By utilizing natural language processing.
- **Key steps performed:** Data Input (Data Source: Interview Transcript), Preprocessing (Text Data Processing, Prompt Construction for GPT), Modeling (Interaction with GPT-3.5 Turbo), Model Evaluation, and Output Generation (Feedback Output and Evaluation).
- Utilized **tools** like libraries: OpenAI API
- Followed **best practices** by following a **structured approach** starting from **reading the transcript to generating and storing the feedback.**
- This makes the system flexible and easy to maintain or expand, Automated Feedback System by using **GPT-3.5 turbo**, this project automates the process of generating structured feedback, reducing the manual effort required to assess candidate performance.
- **Github:** github.com/saikale/Hackathon

5. INSIGHTHIRE – EMPLOYABILITY ANALYTICS PLATFORM

- **Objective:** To bridge graduate employability gaps in Ireland by developing an AI-driven platform that enhances job–candidate matching, improves resume alignment, and provides referral and upskilling insights.

- **Key steps performed:**

1. **Data Acquisition:** Job postings scraped using Octoparse (LinkedIn, Indeed), and resumes parsed with Google Document AI.
2. **Data Integration:** Standardized job titles, skills, and eligibility criteria using Python and manual curation.
3. **Modeling & Analysis:** Developed multiple predictive models including Skill Fit Score (XGBoost), Hiring Trends (ARIMA), Role Progression (Random Forest), and indexes like Risk Prediction & Opportunity Index.
4. **Visualization & Delivery:** Built a user-facing Streamlit app integrated with interactive Power BI dashboards to deliver insights such as skill-fit, referral pathways, and hiring trends.
5. Followed **best practices** by following a **structured approach** starting from **reading the transcript to generating and storing the feedback**.

- **Tools & Technologies:** Python (Pandas, Scikit-learn, XGBoost, NLP), Streamlit, Power BI, ARIMA, Random Forest, Google Document AI, Octoparse.
 - **Professional Learning:** Learned to manage ambiguous datasets, iterate through prototyping, and practice agile documentation. Enhanced problem-framing and stakeholder storytelling skills while ensuring ethical handling of user data.
-