**Project Title**

**A. “Global Trends in Education Affordability: A Data-Driven Analysis of International Tuition Costs”**

**B. “Labor Market Insights from LinkedIn: Predictive Modelling of Job Trends and Skill Demand”**

**2. Research Questions**

**A. Education Dataset:**

* **How do tuition costs vary across countries and fields of study?**
* **What socio-economic indicators correlate with higher or lower education costs?**

**B. LinkedIn Job Dataset:**

* **What are the key skills and qualifications most frequently associated with in-demand jobs?**
* **Can we predict job market trends based on current job postings?**

**3. Hypotheses**

**A. “Global Trends in Education Affordability: A Data-Driven Analysis of International Tuition Costs”**

This idea suggests that countries with stronger economies (measured by GDP per person) are likely to charge more for international students to study there. The reason is that wealthier countries often have better universities, more advanced facilities, and attract more students from around the world—so they can charge higher tuition fees. Also, countries with higher living costs and stronger currencies usually end up having more expensive education.

**How a country’s economy affects the price of studying there as an international student.**

**Whether studying abroad is becoming something only wealthier students can afford.**

**How students and governments can better understand and respond to the rising cost of international education.**

**B. “Labor Market Insights from LinkedIn: Predictive Modelling of Job Trends and Skill Demand”**

Today’s job market is heavily influenced by online platforms like LinkedIn, which collect up-to-date information about what employers are looking for—such as required skills, qualifications, and job locations. This hypothesis suggests that jobs that are in high demand tend to follow clear and consistent patterns. These patterns can be recognized and predicted using data and machine learning.

These patterns usually show up in three main areas:

1. **Skills:**  
   Jobs that are in high demand often ask for the same types of skills over and over. For example, data-related jobs often require skills like Python, while office jobs may ask for project management or communication skills. These skill trends often match what’s happening in the wider world—like the growing use of AI or digital tools.
2. **Industries:**  
   Some industries regularly post more job openings than others. Technology, healthcare, finance, and logistics are good examples. These fields often need more workers because they are growing fast, changing quickly, or facing new rules that require more specialized roles.

By using machine learning, we can find these patterns in job listings and use them to predict which roles will be in high demand, what skills are needed, and where those jobs are likely to be found.

**4. Aims and Objectives**

* **Aim**: To analyse the dataset of international cost of education and LinkedIn job posting to identify key risk factors and develop a predictive model for international education and LinkedIn job posting.
* **Objectives**:
  1. **Data Preprocessing**: Clean and preprocess the dataset, handling missing values and normalizing data where necessary. (Collect and preprocess global education cost data And Scrape/preprocess LinkedIn job data (e.g., using APIs or third-party datasets).
  2. **Exploratory Data Analysis (EDA)**: Conduct EDA to identify patterns and relationships within the dataset using visualizations. (Visualize cost trends by region, field of study, and economic indicators and Use NLP to extract job titles, required skills, and sectors)
  3. **Model Development**: Apply regression models to understand predictors of tuition variability and Apply classification/clustering models to identify job clusters and demand patterns.to determine which best predicts **international education and LinkedIn job posting.**
  4. **Model Evaluation:** Evaluate model performance using RMSE and R² metrics. Evaluate skill-demand models using accuracy, precision, and recall.
  5. **Background:**

This project explores two key global issues: the rising cost of **international education and job posting market trends using LinkedIn data**. It aims to understand how a country’s economic strength affects tuition fees for international students and whether studying abroad is becoming less accessible. At the same time, it analyses job postings to identify which roles, skills, and industries are most in demand. By using data analysis and machine learning, the project seeks to provide insights that can help students make better decisions about education and careers.

**5. Data**

The datasets to be used is the, sourced from the Kaggle website ([**https://www.kaggle.com/datasets?search=international+education+cost**](https://www.kaggle.com/datasets/arshkon/linkedin-job-postings)). It contains **900records** and **12 features** and merging the other data set **LinkedIn job posting** ,This data contains various types of data like ( companies, skills, job posting and industries) **24000 records and 15 features** ([**https://www.kaggle.com/datasets/arshkon/linkedin-job-postings**](https://www.kaggle.com/datasets/arshkon/linkedin-job-postings)) including both continuous (e.g., GDP Country, university etc..)and categorical variables (e.g., tuition fee, course duration ,etc…). Key attributes include clinical measurements and demographic information relevant to education and job posting data.

**5.1 Methodology**

**Step 1: Data Collection**

* **Education**: Use open datasets (e.g., Kaggle dataset, World Bank, OECD).
* **LinkedIn**: Scrape or use publicly available job data aggregators (with ethics clearance).

**Step 2: Data Preprocessing**

* Handle missing values, convert currencies, and normalize values (Education).
* Use NLP techniques (tokenization, lemmatization, stop-word removal) on job descriptions (LinkedIn).

**Step 3: EDA & Visualization**

* Use heatmaps, boxplots, and time series (Education).
* Use word clouds, bar charts, and frequency plots (LinkedIn).

**Step 4: Model Development**

* **Education**: Linear/Multiple Regression, Decision Trees.
* **LinkedIn**: Naive Bayes, Random Forest, K-means (clustering by industry/skill).

**Step 5: Model Evaluation**

* **Education**: RMSE, R².
* **LinkedIn**: Evaluate skill-demand models using accuracy, precision, and recall.

**6. Tools and Technologies**

* **Python** (pandas, NumPy, scikit-learn, matplotlib, seaborn, NLP models)
* **Jupiter Notebooks/google colab**
* **Power BI/Tableau** for dashboards

**7. Data Ethics**

Use data from publicly available sources with proper citation. No personally identifiable information (PII) will be processed. Ensure compliance with GDPR/CCPA if applicable (especially for LinkedIn data). Obtain approval if scraping platforms with TOS restrictions.

**8. Project Plan**

| **Task** | **Description** | **Start Date** | **End Date** |
| --- | --- | --- | --- |
| Literature Review | Review existing research on cost of international education and LinkedIn job posting prediction and modelling | May 30 2025 | June 06 2025 |
| Data Preprocessing | Clean and prepare the dataset for analysis | June 08 2025 | June 13 2025 |
| Exploratory Data Analysis | Perform EDA to identify trends and relationships | June 15 2025 | June 21 2025 |
| Model Development | Implement regression models to analyse the data | June 23 2025 | June 28 2025 |
| Model Evaluation | Assess model performance and refine as needed | July 2 2025 | July 11 2025 |
| Presentation | Presenting off the models and data visualization PPT | July 12 2025 | July 18  2025 |
| Final Report Writing | Compile findings, conclusions, and recommendations | July 20 2025 | July 28 2025 |

**8.1 Data Management Plan**

* **Data Collection**: The dataset will be downloaded from the Kaggle data set and imported into a suitable analysis tool (e.g., pandas, NumPy, scikit-learn, matplotlib, seaborn, NLP models), **Power BI (**for dashboards)
* **Storage**: The dataset and project files will be stored on a local machine and backed up to a cloud service (e.g., Google Drive).
* **Version Control**: GitHub will be used to track changes and manage code versions throughout the project, facilitating collaboration and backup.
* **Backup**: Regular backups will be scheduled weekly to ensure data integrity.

**9. Reference List and Bibliography**

* For International Education Cost Dataset and Background:
* UNESCO. (2020). *Global Education Monitoring Report 2020: Inclusion and education – All means all*.
* OECD. (2021). *Education at a Glance 2021: OECD Indicators*. OECD Publishing.
* For LinkedIn Job Postings Dataset and Background:
* Zhou, Y., Sharma, A., & Budak, C. (2021). *Characterizing Job Market Trends with Online Job Postings*. Proceedings of the International AAAI Conference on Web and Social Media.
* Choudhury, M. D., Sundararajan, V., & Sharma, A. (2019). *Analyzing the LinkedIn Employment Market Using Machine Learning Techniques*. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining.