Project Report

Corporate Scams & Market Research
-EdTech & HR Domain

Submitted by: Data Analytics Team A

Submitted to: Wasim Patwari (CEO and Founder)

Acknowledgement

We would like to express our heartfelt gratitude to all those who guided and supported us throughout the completion of our project titled "**Corporate Scams & Market Research in the EdTech & HR Domain**." First and foremost, we would like to sincerely thank **Mr. Wasim Patwari (CEO and Founder)** for giving us the opportunity to work on this insightful project.

His visionary leadership and encouragement provided us with valuable exposure and direction.

We are especially grateful to our mentor, **Ms. Shravini, Ms. Najuka** for her constant guidance, patience, and motivation throughout the project journey. Her timely feedback, domain expertise, and continuous support were instrumental in shaping the quality of our work.

We also extend our appreciation to **Mr Harshal for leading the entire team** and organization for providing us with a platform to explore real-world business problems and apply analytical and research skills effectively.

This project has been a significant learning experience and would not have been possible without the contributions of everyone mentioned above.

Abstract

This project, titled "Corporate Scams & Market Research in the EdTech and HR Domain," aims to explore the increasing cases of corporate fraud and unethical practices in two fast-growing sectors: **EdTech and HR**. The primary objective was to identify scam patterns, analyze market behavior, and assess the risks that impact investors, companies, and consumers alike.

To achieve this, a combination of online research, **web scraping,** and **data analysis** tools such as Python, SQL, Excel, and Power BI were used to collect, process, and visualize data. The study involved analyzing **real-world fraud cases**, industry trends, and company behavior through publicly available data.

The findings highlight key **red flags, including misleading marketing, false promises, data manipulation, and lack of regulatory oversight.** This project provides actionable insights that can help businesses, stakeholders, and regulators take preventive steps towards building trust and **transparency in the corporate ecosystem.**

INDEX

Introduction

In today's digital world, corporate scams have become a serious issue, especially in sectors like **EdTech and Human Resources** (HR). With the rapid growth of online education platforms and digital hiring processes, these industries are more vulnerable to fraudulent activities. Scams such as fake job offers, false course promises, and misleading advertisements are increasing day by day, affecting thousands of people financially and emotionally.

The EdTech sector has seen a huge rise in startups offering online courses and educational services. Similarly, the HR domain plays a critical role in recruitment and career development. However, both industries are now facing challenges due to increasing cases of scams, lack of proper verification, and growing user trust issues.

This project aims to identify and analyze different types of scams that occur in these sectors, understand their patterns, and evaluate their impact on users and businesses. By using tools like **web scraping**, **Python, SQL, Excel,** and **Power BI,** we collected and analyzed data to uncover scam hotspots, affected age groups, key companies involved, and frequency trends.

The study focuses on scam-related cases in India and provides a data-driven **market research analysis.** The report is structured to include the abstract, objectives, methodology, data findings, visual dashboards, and concluding recommendations to reduce such fraudulent practices in the future.

Objective

The primary objectives of this project are as follows:

- To identify and understand the different types of corporate scams occurring in the EdTech and HR sectors.
- To analyze the impact of these scams on users, companies, and market credibility.
- To collect and study real-world data using data analysis tools such as Python, SQL, Excel, and Power BI.
- To detect scam patterns based on factors such as age group, city, company name, and time period.
- To visualize the data in the form of interactive dashboards and graphs for better understanding and decision-making.
- To highlight the major scam hotspots and target areas where such frauds are most frequent.
- To offer insights and preventive recommendations that can help users and organizations avoid falling victim to such scams.

METHODOLOGY

This project followed a structured data analytics lifecycle to ensure the reliability, accuracy, and depth of insights. Below are the key phases of the methodology:

3.1 **Data Collection**

Data was collected from multiple trusted online sources, including **news websites**, **job portals**, **EdTech platforms**, and consumer complaint forums. Web scraping techniques were applied using Python libraries like **BeautifulSoup** and **Requests** to extract relevant information such as company names, scam types, user complaints, affected regions, and age groups. The data was exported into CSV and Excel formats for initial examination.

3.2 **Data Cleaning and Preparation**

Once the raw data was gathered, SQL was used for cleaning, normalization, and transformation.

- Removing duplicate records

This process ensured data integrity and allowed for efficient querying and analysis.

```
-Which companies have the most scam reports?

SELECT TOP 5 [Company_Name], COUNT(*) AS Scam_Report_Count

FROM Scam

GROUP BY [Company_Name]

    ORDER BY Scam_Report_Count DESC;
```

Company_Name	Scam_Report_Count
BYJU'S	320
ScholarLink India	316
Buttress Technologies Private Limited	313
AV Systems	313
Aspentech Global Solutions	312

```
--What is the distribution of victims by age group

SELECT Age_Group,COUNT(*) AS total_victim FROM Scam

GROUP BY Age_Group

ORDER BY total_victim
```

Results Messages

	Age_Group	total_victim
1	30–45	4868
2	20–30	4996
3	18–25	5061
4	25–35	5075

3.3 KPI Identification

Key Performance Indicators (KPIs) were defined to guide the analysis and visualization efforts.
These KPIs included: Number of scams per company Distribution of scams by sector (EdTech vs HR) Scam frequency by age group and location Common types and impact levels of scams.

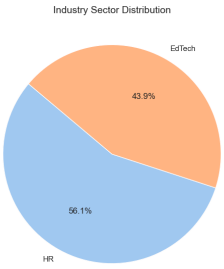
3.4 Data Analysis

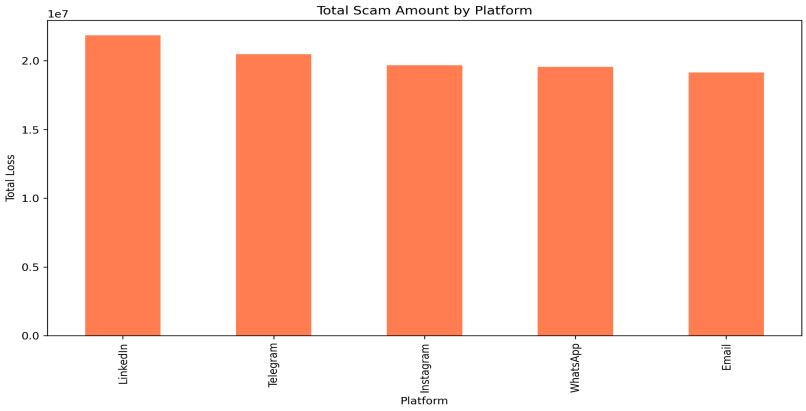
Advanced analysis was conducted using Python libraries such as **pandas, numpy, matplotlib**, and **seaborn**.
This phase focused on: Identifying scam trends over time
Segmenting data based on demographics

Discovering correlations (e.g., between scam type and age group)

Grouping companies based on frequency and type of complaints

These patterns provided **critical insights** into how scams operate and who is most affected.

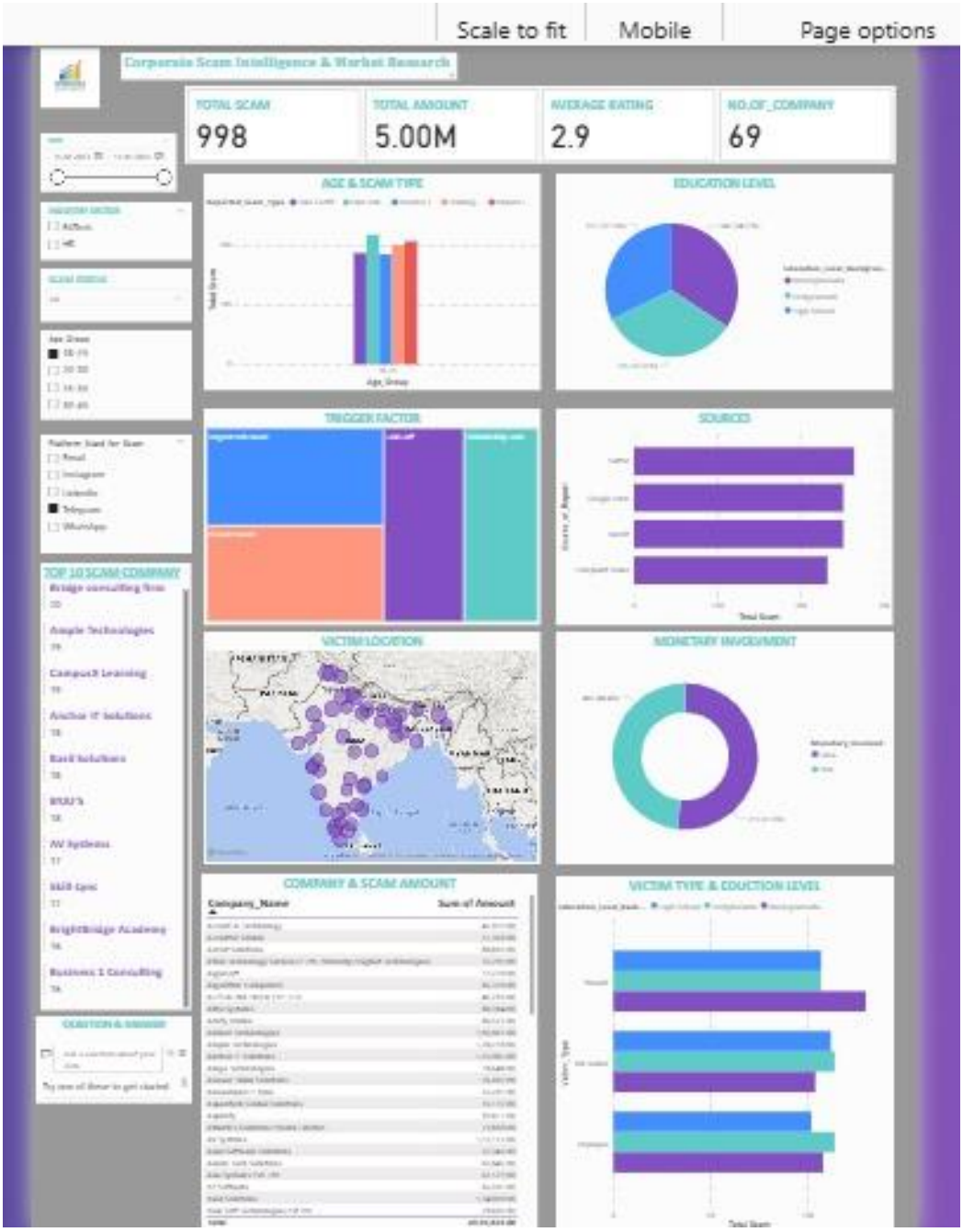




3.5 Data Visualization

Power BI was used to develop interactive dashboards and visual reports.

- **The visualizations included**: Bar and column charts showing scam volumes per region or company.
- **Heatmaps** to highlight geographical scam hotspots.
- **Pie charts** displaying scam type distribution.
- **Slicers** and **filters** for drill-down by sector, age group, and location.
- These visuals made the insights **accessible** and **actionable** for stakeholders.


Scale to fit
Mobile
Page options
Corporate Scam Intelligence & Market Research
TOTAL SCAM
998
TOTAL AMOUNT
5.00M
AVERAGE RATING
2.9
NO.OF_COMPANY
69
AGE & SCAM TYPE
EDUCATION LEVEL
TRIGGER FACTOR
SOURCES
TOP 10 SCAM COMPANY
VICTIM LOCATION
MONETARY INVOLVEMENT
COMPANY & SCAM AMOUNT
VICTIM TYPE & EDUCTION LEVEL

TOOLS & TECHNOLOGY

In this project, various tools and technologies were used at different stages—from collecting raw data to presenting final insights through visual dashboards. Each tool played a specific role in simplifying, analyzing, and communicating the data effectively.

1. Python Programming Language

Python is a high-level, user-friendly programming language known for its powerful capabilities in data analysis, automation, and web scraping.

How it was used in the project:

- To extract scam-related data from websites using web scraping techniques.
- To clean and organize large datasets using libraries like pandas.
- To identify trends and patterns, such as most affected locations, age groups, and companies.
- To create basic visual charts using libraries like matplotlib and seaborn

2. Jupyter Notebook

Jupyter Notebook is an open-source interactive environment where code, output, visualizations, and explanations can be written together in a single file.

- How it was used in the project:
- For writing and testing Python code step-by-step.
- To display data cleaning, analysis, and graphs in one organized place.
- To document the entire process in a readable format for transparency and easy understanding.

3. Python Libraries

Library: Purpose in Project

Requests : Sent HTTP requests to websites and retrieved raw HTML content for scraping.

BeautifulSoup: Parsed HTML pages to extract meaningful data like company names and scam types.

Pandas :Cleaned and transformed the raw data into structured tables (DataFrames).
matplotlib & seaborn: Created graphs to visualize trends, distributions, and comparisons.

4. Microsoft Excel

Excel is a spreadsheet tool widely used for quick data viewing, formatting, and filtering.

How it was used in the project:

- Opened .csv files generated from Python scripts.
- Converted data to .xlsx format for easier compatibility with Power BI.
- Applied filters and formulas for basic summaries and manual validation.
- Prepared the cleaned data before importing it into Power BI or SQL.

5. SQL (Structured Query Language)

SQL is used to manage and query structured data in databases.

How it was used in the project:

- Structured the cleaned data into proper tables.
- Queried the data to find scam frequency by company, age group, and city.
- Helped in filtering and grouping data for analysis and visualization.

6. Power BI

Power BI is a business intelligence tool used to create interactive dashboards and reports.

How it was used in the project:

Imported final data from Excel.

Built visuals like:

- **Bar charts** (e.g., Top scam cities)
- **Pie charts** (e.g., Scam types)
- **Heatmaps** (e.g., Region-wise scam density)

.

IMPLEMENTATION

The implementation phase followed a systematic approach to ensure data was collected, cleaned, analyzed, and visualized effectively. A step-by-step methodology was followed to meet the project objectives, with well-defined responsibilities and the use of advanced tools and technologies.

1. Goal Definition

The primary goal of the project was to analyze corporate scams and conduct market research in the EdTech and HR sectors. The focus was on identifying scam trends, understanding affected demographics, and generating data-driven insights using real-world data sources.

2. Data Collection (Web Scraping)

Web scraping was performed to extract data from relevant and trusted online sources.

Technologies Used: Python, Requests, BeautifulSoup

Responsibilities: Priyanka,Tarun and Harshal

Outcome: Raw data collected in unstructured format (HTML/tables) for further processing.

3. Data Cleaning and Preprocessing

The scraped data was cleaned and converted into a structured format using Python libraries.

Libraries Used: Pandas, NumPy

Tasks Performed: Removal of null values, duplicates, data type conversion, column renaming

Responsibilities: Tharun

Outcome: A clean and analysis-ready dataset in Excel/CSV format.

4. Data Storage and Exploration (Excel)

After cleaning, the data was stored in Excel for manual checks and

initial exploration.

Functions Used: SUM, COUNT, AVERAGE.

Purpose: To verify the accuracy of data and prepare it for Power BI

Outcome: Structured datasets with basic summaries and formulas.

5. Data Analysis (SQL)

Structured Query Language (SQL) was used for deeper querying and insight extraction.

Functions Used: WHERE clauses, aggregate functions like SUM(), COUNT(), AVG()

Purpose: To filter **Windows function** ,**aggregate** data based on location, company, age group, etc.

Responsibilities: Priyanka

Outcome: Efficient extraction of targeted insights from large datasets.

6. Visualization and Dashboarding (Power BI)

Interactive dashboards were built using Power BI to visualize and present insights.

Charts: Bar, Pie, Map, and Line charts

DAX: Calculated columns and measures

Slicers & Filters: For interactive drill-downs

Table Merging and Relationships

Responsibilities:

Dashboard Design: **Priyanka**

Outcome: A professional and user-friendly dashboard highlighting scam hotspots, affected age groups, fraud amounts, and more.

7. Documentation and Reporting

All steps, insights, and visuals were compiled into a comprehensive project report.

Responsibilities: Priti Priya Sahu

Tasks: Drafting the implementation process, summarizing findings, designing layout, and formatting for submission.

CHALLENGES

During the execution of the project, several data-related challenges were encountered which impacted the workflow and required extra attention during the preprocessing phase.

1. Difficulty in Finding Reliable Data

Identifying and collecting accurate data was one of the initial challenges. The data was not available from a single source, so information had to be gathered from multiple websites. Navigating through different platforms to retrieve relevant content was time-consuming.

2. Unstructured Data Formats

The data collected from various sources was not in a uniform structure. It appeared in different formats, such as scattered tables, unformatted text, and embedded HTML elements, which made it difficult to extract and organize efficiently.

3. Presence of Data Errors

The raw dataset contained common issues such as:

Duplicate records,Null entries,Missing values.
These inconsistencies impacted the reliability and required thorough cleaning before analysis.

4. Incorrect Date Format

Another challenge was that date-related fields were stored in string format instead of a proper date datatype. This created problems during sorting, filtering, and performing time-based analysis in tools like Excel, SQL, and Power BI.

These challenges affected the initial stages of the project but were addressed carefully during the data preparation process to ensure smooth analysis and visualization later.

RESULTS /ANALYSIS

1. **Total Scam Reports Analyzed**:
    - **20,000** scam incidents were recorded across various platforms and domains.
    - HR : **11,223**
    - EdTech :**8,777**
2. **Top Companies by Scam Reports**:
    - BYJU'S **(320)**, ScholarLink India **(316)**, and Buttress Technologies Pvt. Ltd. **(313)** had the highest number of scam complaints.
3. **Most Targeted Age Group**:
    - **25–35 years** recorded the highest victim count, followed closely by the **18–25** group.
4. **Most Affected Education Level**:
    - **Undergraduate** background was the most frequently targeted (6,705 cases).
5. **Most Common Scam Types**:
    - **Fake Job Posting**, **Resume Fee Scam**, and **Training-Fee Trap** were the top scam types, with each exceeding **3,900 cases**.
6. **Top Platforms Used for Scams**:
    - **LinkedIn**, **Email**, and **Telegram** were the most used platforms, each involved in over **4,000 scams**.

7. **Monthly Scam Trend**:

    - **October** showed the highest number of scam reports (**1,764 cases**), followed by **December** and **March**.

8. **Geographic Hotspots**:

    - **Mumbai (2361)**, **Chandigarh (2248)**, and **New Delhi (2238)** are the cities with the highest number of scam-generating companies.

    - **Tirupati**, **Kanpur**, and **Noida** reported the most scam victim incidents.

9. **Average Duration of Scam Campaigns**:

    - The average scam campaign lasted approximately **16 weeks**.

    - **Scam Reports with High Risk** (₹10,000+ and credibility rating ≥ 4):

- 30+ such high-severity reports identified, often involving **false job guarantees**, **forced payments**, and **fake certifications**.

SOLUTION

Suggestions & Solutions: How to Create Awareness and Protect Against Corporate Scams

To reduce the risk and impact of corporate scams in EdTech and HR sectors, the following solutions and awareness measures are recommended:

1. Awareness Campaigns

- Organize **webinars, workshops**, and **social media** campaigns to educate job seekers and learners about common scam techniques.
- Collaborate with **colleges and placement** cells to reach fresh graduates.

2. Verification of Sources

- Always verify the official **website**, **email domain**, and **contact details** before applying or making any payments.
- Use platforms like **LinkedIn**, Glassdoor, and **government-registered** job portals to confirm legitimacy.

3. Avoid Upfront Payments

- Genuine companies do not demand **money for job applications,** training, or interviews.
- Be cautious of any organization asking for **registration fees**, **deposits**, or investment promises.

4. Use Secure Communication Channels

- Avoid sharing **sensitive documents** (like Aadhaar, PAN, or bank details) over WhatsApp or Telegram.
- Ensure emails are coming from **verified domains** (e.g., @companyname.com instead of @gmail.com).

5. Report Suspicious Activities

- If you encounter a scam, report it immediately to **cybercrime portals** (https://cybercrime.gov.in) or local authorities.
- Inform others via online **reviews** to prevent further victimization.

6. Build Digital Literacy

- Teach individuals how to identify **phishing emails,** fake job offers, and fraudulent course schemes.
- Encourage critical thinking and **fact-checking** before **trusting** online offers.

7. Strong Policies and Employer Audits

- EdTech platforms and HR consultancies should undergo regular audits and **background** checks.

- Government and industry **bodies** should enforce stricter **compliance** and penalties for fraud.

CONCLUSION

This project successfully explored the patterns and impacts of corporate scams within the **EdTech** and **HR** sectors by collecting, analyzing, and visualizing **real-world data.** Through structured implementation—including **data collection**, **cleaning**, **analysis,** and **dashboard creation**—we identified key trends such as the most affected **regions**, common scam methods, and targeted **age groups**.

The findings highlight the urgent need for **awareness, transparency**, and proper verification practices among **job seekers** and **students**. By leveraging tools like **Python**, **Excel**, and Power BI, we were able to transform raw data into **actionable** insights. These insights can support individuals, **organizations**, and **policymakers** in developing better safeguards against fraudulent activities and improving trust in **digital hiring** and learning environments.

REFERENCE

Fake company links

- https://careerswami.com/fake-companies-list/
- https://wallisinfo.com/fake-companies-list
- https://www.consumercomplaints.in/airports-authority-of-india-b115912

THANK YOU