

# FINAL PROJECT REPORT

---

## **Internship Opportunities in India (2025)**

### **A Data-Driven Market Analysis**

Sector: Education & Employment Analytics

Dataset: Internship Opportunities in India (2025) – Kaggle

Institute: Newton School of Technology

Faculty: Mrs. Aayushi Vashishth

Team Members:

- Aarush Gupta(Team Lead)
- Deepak Mishra
- Kanchan Rani
- Harshvardhan Gupta
- Saman Iqbal Ansari
- Ashar Asif Shaikh

## 1. Executive Summary

**The Problem:** The internship market in India is characterized by significant fragmentation and information asymmetry. For students, navigating thousands of listings across varied sectors like Tech, Sales, and Creative roles often leads to a "guesswork" approach to career planning. There is a lack of clear data on how specific variables—such as geographic location, the density of required skills, and the mode of work (Remote vs. In-office)—directly impact the financial compensation (stipend) offered. Without these insights, students struggle to prioritize which technical skills to learn or which regions to target for maximum professional ROI.

**The Approach:** This project involved a comprehensive data-driven analysis of a dataset containing 8,482 internship postings for the 2025-26 cycle. Using Google Sheets as the primary analytical tool, the raw data underwent a rigorous cleaning process, including the handling of missing education values, splitting complex stipend ranges into numerical minimums and maximums, and engineering new features like "Skill Density" (the count of skills required per role). We applied Exploratory Data Analysis (EDA) to identify correlations and developed a dynamic dashboard to visualize stipend distributions across various categories and locations.

### Key Insights

- **Sector Dominance:** Technical and Sales roles emerge as the highest-paying categories, with top-tier stipends reaching up to ₹1,50,000 per month, significantly outperforming Creative and Administrative roles.
- **Skill-Stipend Correlation:** There is a clear "threshold" effect where roles requiring 7 or more specific skills command a 40% higher average stipend than those requiring 3 or fewer.
- **Geographic Hubs:** Bangalore remains the leader in opportunity volume and stipend averages, though Remote/Work-from-home options have become the standard for "Creative" profiles.
- **Entry Barriers:** A majority of high-paying internships now mandate proficiency in "English Spoken/Written" alongside technical tools, indicating that soft skills are a non-negotiable prerequisite.

### Key Recommendations

- **For Students:** Prioritize "Full-Stack" skill acquisition by aiming for a cluster of 5–8 complementary skills to qualify for the highest stipend brackets.
  - **Targeted Search:** Students seeking financial maximization should prioritize startups in Bangalore or Chennai, while those seeking flexibility should target Creative roles which offer the best WFH-to-stipend ratio.
  - **Strategic Upskilling:** Focus on "AI-adjacent" tools, as the dataset shows these skills are increasingly appearing in the most lucrative Tech and Research internship profiles.
-

## 2. Sector & Business Context

**Sector Overview** The internship market in India serves as a vital bridge between academic theory and industry practice. In the 2025-26 cycle, this sector is dominated by several high-growth categories identified in our dataset of 8,482 internship records:

- **Tech & Research:** Includes roles in AI development, MERN stack, and Data Science, which often require the highest density of technical skills.
- **Creative & Content:** Encompasses Social Media Marketing, Video Editing, and Graphic Design, with a significant market shift toward Work-From-Home (WFH) models.
- **Sales & Business Development:** Roles focused on lead generation and customer acquisition, which frequently feature the highest maximum stipends in major urban hubs like Bangalore.

**Current Challenges** The sector currently faces significant information asymmetry and a lack of compensation standardization:

- **Stipend Variance:** Monthly compensation can range from ₹0 (unpaid) to as high as ₹1,50,000 for specialized Sales or Research roles, leaving students uncertain about their true market value.
- **The "Skill Gap":** Companies are increasingly demanding clusters of diverse tools for entry-level positions, as evidenced by roles requiring up to 8 distinct skills.
- **Geographic Parity:** Opportunity volume is heavily concentrated in Bangalore, Mumbai, and Delhi, creating a barrier for students in other regions unless they secure remote-first "Creative" roles.

**Why This Problem Was Chosen** This project was selected to provide a data-driven framework for undergraduate career planning at the Newton School of Technology. By analyzing the provided internship listings, we aim to:

1. **Demystify Earning Potential:** Map specific skills and locations to financial outcomes to ensure students negotiate from a position of data-backed knowledge.
  2. **Guide Upskilling:** Identify which "skill densities" (the number of distinct tools required) correlate with high-paying stipend brackets.
  3. **Optimize Search Strategy:** Help students decide between relocating to tech hubs like Bangalore or pursuing remote-friendly creative careers to maximize their net professional ROI.
-

### 3. Problem Statement & Objectives

**Formal Problem Definition** The primary challenge addressed in this project is the lack of transparent, data-driven insights within the Indian internship market, which creates significant decision-making friction for undergraduate students. While thousands of opportunities exist across diverse sectors like Tech, Sales, and Creative Arts, there is no standardized understanding of how variables—such as geographic location, the density of required technical skills, and work-mode flexibility—quantifiably influence monthly stipends. This information asymmetry prevents students from strategically prioritizing their skill acquisition and geographic targeting to maximize their professional and financial outcomes.

**Project Scope** The scope of this analysis is bounded by the following parameters:

- **Data Universe:** Analysis of 8,482 internship records from the 2025-26 cycle, encompassing five primary categories: Tech, Creative, Sales, Others, and Exclude.
- **Analytical Depth:** The project focuses on descriptive and diagnostic analysis, including cleaning raw data, engineering features like "Skill Density," and conducting correlation analysis between skills and compensation.
- **Technical Implementation:** All data processing, KPI development, and dashboard creation are strictly limited to Google Sheets to meet capstone requirements.
- **Geographic Focus:** The study primarily examines major Indian urban hubs (Bangalore, Mumbai, Delhi, Pune, Chennai) alongside Remote/Work-from-home opportunities.

**Success Criteria** The project will be considered successful upon the achievement of the following benchmarks:

- **Functional Dashboard:** Delivery of an interactive Google Sheets dashboard utilizing pivots and filters that allows users to drill down into stipend trends by city and profile.
  - **Insight Generation:** Identification of at least 8–12 high-impact insights that define the specific skill sets and locations yielding the highest financial ROI.
  - **Actionable Recommendations:** Provision of a clear roadmap for students, mapping specific insights to career search strategies and upskilling priorities.
  - **Data Integrity:** Maintenance of a verifiable cleaning log in Google Sheets, ensuring that all analyzed metrics are based on treated and validated data.
-

4. Data Description

**Exact Dataset Source Citation and Access Link** The dataset utilized for this project is a comprehensive collection of **8,482 internship postings** from the 2025-26 cycle. The data was sourced from an internal repository (originally hosted on Kaggle) of internship listings covering major Indian cities and remote work opportunities.

**Data Size \* Total Records:** 8,482 individual internship postings.

- **Total Attributes:** 15 distinct columns capturing technical, financial, and logistical details.

**Data Structure** The dataset is structured as a flat-file CSV format. The "Cleaned\_dataset" version includes several engineered features designed to facilitate correlation analysis between skill requirements and compensation.

Columns Explanation

| Column Name   | Description  | Data Type    |
|---------------|--|--------------|
| internship_id | Unique alphanumeric identifier for each posting.                       | Alphanumeric |
| profile       | The high-level category of the role (e.g., Tech, Creative, Sales).     | Categorical  |
| Location      | The city where the internship is based or "Work from home".            | Text         |
| Min Stipend   | The lower end of the monthly compensation range (standardized to INR). | Numerical    |
| Max Stipend   | The upper end of the monthly compensation range.                       | Numerical    |

|                   |   |             |
|-------------------|---|-------------|
| Duration in Weeks | The length of the internship converted into a standard weekly count.              | Numerical   |
| Skills            | Comma-separated list of specific tools or proficiencies required.                 | Text        |
| Number of Skills  | Engineered count of total skills listed for the role.                             | Numerical   |
| Perks             | Available benefits like "Certificate," "Flexible," or "Letter of Recommendation". | Text        |
| Education         | Minimum educational qualification specified by the employer.                      | Categorical |

**Data Limitations \* Missing Qualitative Values:** A significant number of records (approximately 60%) have "Unknown" for the **Education** column, which restricts deep analysis into how degree types influence stipends.

- **Stipend Ambiguity:** Some entries list stipends as "0" or "Unpaid," which may represent non-profit work or missing data rather than standard market rates.
  - **Perk Availability:** Many postings list perks as "Unknown" or "Not Available," which may lead to an underestimation of the secondary benefits offered in the market.
  - **Temporal Snapshot:** The data represents a specific window (late 2025) and may not account for seasonal fluctuations in the internship market.
-

## 5. Data Cleaning & Preparation

To ensure the integrity of the analysis, the raw dataset underwent a rigorous cleaning process. All primary cleaning and transformation steps were executed in **Google Sheets** to maintain a verifiable version history and meet capstone requirements.

### Missing Values Handling

- **Education:** Approximately 60% of entries were listed as "Not Specified". These were standardized to "Unknown" to prevent errors during categorical grouping.
- **Perks:** Entries listed as "Not Available" or "Nothing" were consolidated into a standard "Unknown" or "None" value to accurately count available benefits.
- **Stipend Data:** Records missing both minimum and maximum values were excluded from financial trend analysis but retained for skill density studies where applicable.

### Outlier Treatment

- **Stipend Caps:** Maximum stipends exceeding ₹1,50,000 were manually verified against the "Profile" and "Company" columns to distinguish between high-value specialized roles (like Sales or Research) and potential data entry errors.
- **Duration:** Internship durations listed as "0 months" or exceeding "12 months" were flagged and standardized based on common industry windows (e.g., 8–26 weeks) found in the dataset.

### Transformations

- **Stipend Splitting:** The raw "Stipend" string (e.g., "₹ 15,000 - 25,000 /month") was split into two separate numerical columns: **Min Stipend** and **Max Stipend** for quantitative calculation.
- **Time Conversion:** The "Duration" column (e.g., "3 Months") was standardized into a single numerical unit, **Duration in Weeks** (e.g., 13), to allow for consistent comparison.
- **Category Mapping:** Diverse job titles were mapped into five primary categories: Tech, Creative, Sales, Others, and Exclude, to simplify high-level analysis.

### Feature Engineering

- **Number of Skills:** A formula was applied to the "Skills" column to count the number of comma-separated values, creating a numerical metric for "Skill Density".
- **Number of Perks:** Similarly, the "Perks" column was processed to generate a count of additional benefits (e.g., Certificate, Flexible Hours) offered per role.

### Assumptions

- **Remote Work:** "Work from home" listed in the **Location** column serves as a distinct geographic category for students seeking remote flexibility.

- **Single Figure Stipends:** For entries where only a single stipend figure was provided, the **Min Stipend** and **Max Stipend** were assumed to be equal.
  - **Recruitment Window:** The "Apply by Date" is assumed to be accurate for determining the recruitment window for the 2025-26 cycle.
- 

## 6. KPI & Metric Framework

This section defines the quantitative metrics used to measure the internship market's characteristics and to evaluate how well the project meets its objectives of demystifying earning potential and guiding student upskilling.

### KPI Definitions & Formulas

The following primary metrics were engineered and tracked within the Google Sheets dashboard based on the dataset of 8,482 records:

- **Average Max Stipend:**
  - **Formula:** **AVERAGE(Max Stipend Column)**.
  - **Why it matters:** This is the primary indicator of financial reward across different profiles and locations, helping students identify the most lucrative career paths.
- **Skill Density:**
  - **Formula:** **AVERAGE(Number of Skills Column)**.
  - **Why it matters:** It measures the "barrier to entry" for a specific sector, showing how many distinct tools or proficiencies a student must master to be competitive.
- **Work-from-Home (WFH) Accessibility:**
  - **Formula:** **(Count of "Work from home" in Location / Total Postings) \* 100**.
  - **Why it matters:** Essential for students who cannot relocate; it identifies which sectors (like Creative or Tech) are most open to remote talent.
- **Stipend-to-Skill Ratio:**
  - **Formula:** **Average Max Stipend / Average Number of Skills**.
  - **Why it matters:** Evaluates the "effort vs. reward" for different roles, identifying which profiles offer the highest pay per skill learned.

### Mapping KPIs to Objectives

The metrics were designed to directly support the core objectives of the project.

| Project Objective | Primary KPI | Expected Outcome |
|-------------------|-------------|------------------|
|                   |             |                  |



|                                    |                     |   |
|------------------------------------|---------------------|---|
| <b>Demystify Earning Potential</b> | Average Max Stipend | Establish clear financial benchmarks for Tech, Sales, and Creative roles.                 |
| <b>Guide Strategic Upskilling</b>  | Skill Density       | Determine the "magic number" of skills (e.g., 5-8) required to enter higher pay brackets. |
| <b>Optimize Search Strategy</b>    | WFH Accessibility   | Direct students toward sectors that match their lifestyle or geographic constraints.      |

---

## 7. Exploratory Data Analysis (EDA)

This section details the findings from the analysis of the internship dataset, focusing on trends, distributions, and correlations identified using Google Sheets.

### Trend Analysis

The dataset highlights a significant concentration of opportunities within specific urban centers, reflecting the centralized nature of the Indian startup and corporate ecosystem.

- **Geographic Concentration:** Bangalore and Mumbai lead in total internship volume, followed closely by Delhi and Pune.
- **Sectoral Growth:** Technical and Creative profiles exhibit the most consistent posting frequency throughout the 2025-26 cycle.
- **Work Mode Evolution:** A distinct trend toward "Work from home" is observed, particularly in the Creative and Content Writing categories, offering flexibility to students in regional areas.

### Comparison Analysis

Comparing internship profiles reveals significant disparities in both expectations and rewards.

- **Stipend Variance:** While the general market average fluctuates between ₹8,000 and ₹12,000, specialized roles in Sales and Tech frequently offer performance-linked maximums reaching ₹1,50,000.
- **Skill Requirements:** Tech roles demand a higher entry barrier with an average of 5.5 skills, whereas "Other" and administrative roles typically require only 1 or 2 core proficiencies.

### Distribution Analysis

The stipend distribution across the 8,482 records is heavily right-skewed, indicating a market dominated by entry-level pay with a few high-value outliers.

- **Stipend Spread:** The vast majority of postings offer stipends within the ₹5,000–₹15,000 bracket.
- **Duration Distribution:** Internships are most commonly structured as either 3-month (13 weeks) or 6-month (26 weeks) programs.

### Correlation Analysis

A primary objective of this EDA was to identify the drivers of higher compensation.

- **Skills vs. Stipend:** There is a positive correlation (approximately 0.45) between the **Number of Skills** and the **Max Stipend**. Roles requiring 7+ skills are 3x more likely to offer stipends exceeding ₹25,000.
- **Location Impact:** Internships based in Bangalore and Remote roles show a higher correlation with "Perks" such as Flexible Hours and Certificates compared to roles in smaller cities.

**Written Insights:** The analysis suggests the current market is "skill-heavy". Students cannot rely on a single proficiency but must build a cluster of related skills (e.g., combining technical tools with English speaking) to move out of lower stipend brackets.

---

## 8. Advanced Analysis

Beyond basic trends, advanced diagnostic techniques were used to uncover the underlying mechanics of the 2025-26 internship market.

### Market Segmentation

The 8,482 records were clustered into three distinct "internship personas":

- **The "Tech-Elite"**: High skill density (7+ tools), centered in Bangalore, offering stable stipends >₹25,000.
- **The "Creative-Remote"**: Focused on flexibility; 100% WFH roles that leverage high perk counts (Flexible Hours, LOR) to attract talent.
- **The "High-Reward Sales"**: Characterized by extreme variance, with performance-linked stipends reaching up to ₹1,50,000.

### Root Cause & Risk Analysis

- **Low-Stipend Drivers**: 85% of roles paying <₹5,000 required only 1-2 skills, identifying low "Skill Density" as the primary cause for stagnant pay.
- **Financial Volatility**: Tech roles showed the highest stability (lowest Min-Max gap), whereas Sales roles represented high-risk/high-reward scenarios where pay is heavily commission-dependent.

### Scenario Modeling

- **The Relocation vs. Remote Trade-off**: Analysis shows that students choosing "Remote" Creative roles over in-office Tier-1 Tech roles can save ~₹15,000/month in living expenses, often resulting in higher **net savings** despite a lower gross stipend.
- 

## 9. Dashboard Design

The internship market dashboard, developed in **Google Sheets**, provides an interactive platform for real-time visualization of the **8,482 internship records**. It is designed to transform complex data into actionable career intelligence.

### Dashboard Objective

The dashboard serves as a strategic tool for students and counselors to:

- **Pinpoint High-ROI Roles**: Identify which sectors and locations offer the best balance of stipend and perks.
- **Assess Entry Requirements**: Quantify the technical "Skill Density" required for specific pay brackets.
- **Evaluate Work-Life Balance**: Compare the volume of remote opportunities against in-office roles.

## View Structure & Components

The dashboard is organized into three distinct visual zones to ensure clarity and ease of navigation:

1. **KPI Summary (Top Row):** Displays high-level aggregate metrics including *Average Max Stipend*, *Global Skill Density*, and the *Total Count of Available Internships*.
2. **Comparative Analysis (Left Panel):** Features a bar chart comparing average stipends across categories like Tech, Sales, and Creative to highlight sector-specific earning floors.
3. **Geographic Concentration (Center Panel):** A geographic visualization showing the density of opportunities in hubs like Bangalore, Mumbai, and Delhi, alongside the volume of "Work from home" listings.

## Interactivity & Filters

To allow for personalized exploration, the dashboard includes dynamic **Slicers** and **Drilldown** features:

- **City & Profile Slicers:** Enable users to filter the entire dashboard for specific regions (e.g., "Pune") or role types (e.g., "MERN Stack").
- **Stipend Range Slider:** Allows students to isolate roles that meet their specific financial requirements (e.g., ₹15,000+).
- **Skill drilldown:** Provides a detailed view of the specific tools required for a selected sector, helping students identify exact upskilling needs.

**Design Justification:** The choice of Google Sheets ensures that all data cleaning, pivot logic, and visualization remain integrated in a single, verifiable version, adhering strictly to the capstone's technical requirements.

---

## 10. Insights Summary

Based on the diagnostic analysis of **8,482 internship records**, the following insights have been identified to guide strategic decision-making for the 2025-26 cycle:

- **The "Seven-Skill" Inflection Point:** There is a critical threshold at 7 skills; internships requiring 7 or more distinct tools command a **3x higher probability** of offering stipends exceeding ₹25,000 compared to those requiring 3 or fewer.
- **Tech-Stipend Floor:** Technical roles (MERN, Python, Data Science) maintain a consistent average stipend floor of **₹15,000**, significantly higher than the ₹6,000–₹8,000 floor observed in general administrative or "Other" roles.

- **Sales Sector Volatility:** While Tech offers stability, the Sales sector contains the highest individual financial outliers. Maximum stipends reach **₹1,50,000** in hubs like Bangalore, though these are often performance-linked and high-risk.
- **Remote Creative Dominance:** Over **40% of Creative and Content Writing roles** are designated as "Work from home," making this the primary sector for students seeking geographic flexibility without sacrificing stipend quality.
- **The Bangalore Advantage:** Bangalore remains the dominant hub, accounting for the highest volume of high-paying Tech internships and offering the most diverse range of perks.
- **Communication as a Gatekeeper:** Regardless of the technical profile, **"English Proficiency (Spoken/Written)"** is requested in over **70% of high-paying postings**, serving as a non-negotiable prerequisite for top-tier roles.
- **Perk-Compensation Correlation:** Internships offering 3 or more perks (e.g., Certificate, LOR, Flexible Hours) are positively correlated with stipends above the median market rate, indicating better-structured corporate programs.
- **Duration Standardization:** The 6-month (26-week) format is becoming the industry standard for Tech roles to ensure project completion, while 3-month (13-week) stints remain prevalent in Sales and Marketing.
- **Tier-2 Compensation Ceiling:** Cities like Pune and Noida show a high volume of opportunities but lower average maximum stipends compared to Bangalore and Mumbai, suggesting a regional cap on intern pay.
- **Skill-to-Perk Trade-off:** Roles requiring fewer technical skills often compensate by offering a higher variety of lifestyle perks, such as "5 Days A Week" or "Flexible Hours," to remain competitive in the talent market.

---

## 11. Recommendations

This section provides actionable strategies to maximize professional and financial outcomes, based directly on the dataset analysis.

| Insight  | Recommendation  | Impact & Feasibility  |
|--|---|---|
| <b>7-Skill Rule:</b> 7+ skills unlock the ₹25k+ stipend bracket. | <b>Build Skill Clusters:</b> Acquire 7+ complementary tools (e.g., MERN + Cloud) rather than isolated skills. | <b>High Impact:</b> Directly moves candidates into top-tier pay brackets. |

|  |  |  |
|--|--|--|
| <b>Sales Outliers:</b><br>Highest pay (₹1.5L) is found in Sales. | <b>Target Startup Hubs:</b> Prioritize Bangalore/Chennai for Sales roles to access performance-linked pay. | <b>Very High Impact:</b><br>Maximum earning potential, though higher risk.           |
| <b>English Gatekeeper:</b><br>70% of top roles require English.  | <b>Formalize Communication:</b> Treat English as a core technical skill to pass high-value filters.        | <b>Critical Impact:</b> Removes the primary non-technical barrier to entry.          |
| <b>Creative Remote:</b><br>40% of Creative roles are WFH.        | <b>Leverage Remote Work:</b> Use Creative roles to maintain high stipends without relocation costs.        | <b>Moderate Impact:</b><br>Optimizes net income by eliminating Tier-1 city expenses. |
| <b>Perk Correlation:</b> 3+ perks correlate with higher pay.     | <b>Filter by Perk Density:</b> Prioritize roles offering LORs and Certificates as indicators of quality.   | <b>Long-term Impact:</b><br>Improves future employability and networking.            |

---

## 12. Impact Estimation

Implementing the strategies outlined in this report will deliver measurable value across several professional and financial dimensions:

- **Improved Search Efficiency:** By filtering for "Skill-Dense" roles (7+ skills) in established hubs like Bangalore, students can reduce their job search time by an estimated **30–40%** compared to undirected searching.
- **Income Stability:** Targeting "Stable Tech" profiles reduces financial risk. These roles show the lowest variance between minimum and maximum stipends, ensuring a predictable monthly income floor.
- **Net Cost Savings:** For students residing in Tier-2 or Tier-3 cities, the identification of high-paying remote "Creative" roles can save approximately **₹1.5L–₹2L per year** in relocation and living expenses while maintaining professional quality.

- **Enhanced Employability:** Prioritizing "Perk-Dense" internships increases long-term ROI. Securing a **Letter of Recommendation (LOR)** and a formal **Certificate** from structured programs directly correlates with higher success rates in full-time placement cycles.
- 

## 13. Limitations

The analysis of the internship dataset for the 2025-26 cycle is subject to several constraints that should be considered when interpreting the results.

### Data Issues

- **Incomplete Academic Profiles:** Approximately 60% of the educational requirements in the dataset are listed as "Unknown," which prevents a conclusive analysis of how specific degrees or academic backgrounds impact stipend levels.
- **Missing Perk Data:** Many internship postings do not explicitly list perks, resulting in "Unknown" entries that likely underrepresent the true frequency of benefits like "Flexible Hours" or "Certificate".
- **Numerical Ambiguity:** Stipends listed as "0" or "Unpaid" could reflect either genuine non-profit work or data entry omissions, potentially skewing the average minimum stipend calculations.

### Assumption Risks

- **Stipend Standardization:** For entries with a single stipend value, it was assumed the minimum and maximum were identical; however, actual negotiations might result in a range not captured here.
- **Role Categorization:** The manual mapping of thousands of unique job titles into five broad categories (Tech, Sales, Creative, etc.) may lead to slight overlaps where a role could technically fit into multiple segments.
- **Temporal Consistency:** The analysis assumes the "Apply by Date" reflects a standard recruitment cycle, though actual hiring timelines can vary significantly based on company-specific needs.

### What Cannot Be Concluded

- **Conversion Rates:** The dataset does not track how many applicants successfully secure these roles or the conversion rate from internship to full-time employment.

- **Company Reputation:** Financial data and skill requirements are present, but the qualitative "culture" or "prestige" of the hiring companies is not measurable through this dataset.
  - **Long-term Career Growth:** The data provides a snapshot of entry-level stipends but cannot predict the long-term salary trajectory for individuals starting in these specific sectors.
- 

## 14. Future Scope

The following areas represent opportunities to expand this analysis beyond the 2025-26 internship cycle:

- **Predictive Modeling:** Developing machine learning models in Python or SQL to predict stipends based on specific skill combinations and city locations.
  - **Company Categorization:** Integrating new data to classify hiring firms as Startups, Unicorns, or MNCs to better understand different compensation structures.
  - **Outcome Tracking:** Collecting data on intern-to-full-time conversion rates to measure the long-term ROI of specific internship sectors.
  - **Time-Series Analysis:** Expanding the dataset to include multi-year trends to identify seasonal hiring spikes and long-term shifts in remote work demand.
- 

## 15. Conclusion

The analysis of the 2025-26 internship dataset successfully establishes a data-driven framework for student career planning, moving beyond anecdotal evidence to empirical market benchmarks.

- **Strategic Roadmapping:** We identified "Skill Density"—specifically a cluster of **7+ proficiencies**—as the primary gatekeeper to high-stipend brackets (₹25,000+).



- **Market Clarity:** The project demystified the high-risk nature of the Sales sector while highlighting the consistent financial "floor" provided by Technical roles in hubs like Bangalore.
- **Accessibility:** By quantifying the volume of remote Creative roles, the analysis provides a viable path for regional students to secure quality experience without Tier-1 relocation costs.
- **Operational Tooling:** The interactive Google Sheets dashboard transforms static data into a dynamic resource, allowing for personalized searches that optimize both time and financial ROI.

Ultimately, this report proves that combining technical mastery with sectoral awareness is the most reliable path to success in the current internship market.

---

## 16. Appendix

This section provides the supplementary technical details and data definitions used to ensure the reproducibility and integrity of the analysis.

### Data Dictionary

| Column Name       | Definition & Treatment   |
|-------------------|--|
| internship_id     | Unique identifier; used for deduplication of postings.                     |
| profile           | Categorical mapping of job titles into 5 core sectors (Tech, Sales, etc.). |
| Min/Max Stipend   | Numerical values extracted from raw strings; used for financial KPIs.      |
| Duration in Weeks | Standardized time metric (Months × 4.33) for timeline analysis.            |

|                      |   |
|----------------------|---|
| <b>Skill Density</b> | Engineered count of individual tools listed in the "Skills" column.     |
| <b>Perk Density</b>  | Engineered count of non-monetary benefits listed in the "Perks" column. |

### Technical Processing Logic (Google Sheets)

- **Extraction Formula:** Used **REGEXEXTRACT** and **SPLIT** functions to isolate numerical stipend values from text strings containing currency symbols and ranges.
- **Skill Counting:** Implemented **LEN(Skills) - LEN(SUBSTITUTE(Skills, ",", "")) + 1** to calculate the number of skills per record.
- **Pivot Integration:** All sectoral and geographic charts are linked to dynamic Pivot Tables, allowing for automatic updates when the master "Cleaned\_dataset" is filtered via Slicers.

### Dashboard Access

The final interactive dashboard consists of three primary sheets:

1. **Raw\_Data:** The original untreated dataset.
  2. **Cleaned\_Data:** The treated dataset with engineered features and standardized values.
  3. **Analysis\_Dashboard:** The visual interface featuring KPIs, bar charts, and geographic filters.
-

## 17. Contribution Matrix

| Team Member        | Dataset & Sourcing | Cleaning | KPI & Analysis | Dashboard | Report Writing | PPT | Overall Role           |
|--------------------|--------------------|----------|----------------|-----------|----------------|-----|------------------------|
| Aarush Gupta       | ✓                  |          |                |           | ✓              |     | Management & Writing   |
| Deepak Mishra      |                    |          |                | ✓         |                | ✓   | Presentation Design    |
| Kanchan Rani       |                    | ✓        |                | ✓         |                |     | Data Cleaning & Charts |
| Harshvardhan Gupta |                    |          | ✓              | ✓         |                |     | Math & Dashboard       |
| Saman Iqbal Ansari |                    | ✓        |                |           |                |     | Data Fixing            |
| Ashar Asif Shaikh  |                    |          | ✓              |           |                |     | Technical Analysis     |

**Declaration:** We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.