

Engineering Graduate Salary Dashboard

CEN445 Introduction to Data Visualization - Assignment Report

İbrahim Ümit Taş, Emre Karaman, Abdurrahman Gülmez

November 2025

1. Dataset Information

- **Name:** [Engineering Graduate Salary Dataset from Kaggle](#)
- **Size:** 2,998 students with multiple features including academic scores (10th/12th/College GPA), technical skills, domain knowledge, college tier, location, specialization, and salary.
- **Column Counts:** 34 columns that contain variables such as Specialization and CollegeTier.

2. Analysis Goals

- Explore relationships between academic performance and salary outcomes
- Analyze the impact of college tier and location on graduate employment
- Identify key factors influencing engineering graduate salaries and visualize educational pathways

3. Visualization Techniques (9 Advanced Charts)

İbrahim Ümit Taş:

- Parallel Coordinates Plot for multi-dimensional analysis,
- Sankey Diagram for flow visualization (Degree→Specialization→College Tier→Salary),
- India Choropleth Map for geographic salary distribution.

Emre Karaman:

- Clustered Correlation Heatmap with hierarchical clustering,
- K-Means Clustering with PCA for profile segmentation,
- Enhanced Bar Chart for specialization distribution.

Abdurrahman Gülmez:

- Personality Box Plot & Radar Chart for trait filtering,
- Treemap,
- 360° Radar Chart for candidate analysis.

Interactive Features: Dropdown menus, sliders, multiselect filters, hover tooltips, zoom, pan, and PNG download capabilities on all charts

4. Key Insights & Findings

- Strong positive correlation between academic performance (GPA) and salary outcomes
- Observable gender disparity in salary distribution across states
- Geographic variation: Metropolitan areas (Maharashtra, Karnataka, Delhi) show higher salary concentrations
- Clustering analysis reveals 3-4 distinct graduate profiles with varying career trajectories
- College tier significantly impacts salaries; Tier 1 colleges show 30-40% higher averages
- Computer Science and Electronics graduates command premium salaries
- Technical domain knowledge and employability test scores are strong predictors of success

5. Technical Implementation

Framework: Streamlit (Python web framework).

Libraries: Plotly (interactive visualizations), Pandas (data processing), Scikit-learn (K-Means clustering, PCA, RandomForest).

Architecture: Multi-page app with landing page and 9 dedicated visualization pages. Data preprocessing handled missing values and derived categorical features.

6. Team Contributions

- **İbrahim Ümit Taş:** Charts 1-3 (Parallel Coordinates, Sankey, India Map), Streamlit integration, deployment, landing page
- **Emre Karaman:** Charts 4-6 (Correlation Heatmap, K-Means Clustering, Bar Chart), visualization design
- **Abdurrahman Gülmez:** Charts 7-9 (Box Plot, Treemap, Radar Chart), data preprocessing and cleaning
- **Report Writing:** Collaborative effort by all team members

7. Repository & Links

- [GitHub Repository](#)
- [Streamlit Dashboard](#)