Techpoint Ai Work Sample

Given the data set, do a quick exploratory data analysis to get a feel for the distributions and biases of the data. Report any visualizations and findings used and suggest any other impactful business use cases for that data.

```
In [6]: # Distribution of 'University' variable
        plt.figure(figsize=(12, 5))
        data['University'].value_counts().plot(kind='bar')
        plt.title('Distribution of University')
        plt.xlabel('University')
        plt.ylabel('Count')
        plt.xticks(rotation=45)
        plt.show()
```


Distribution of University
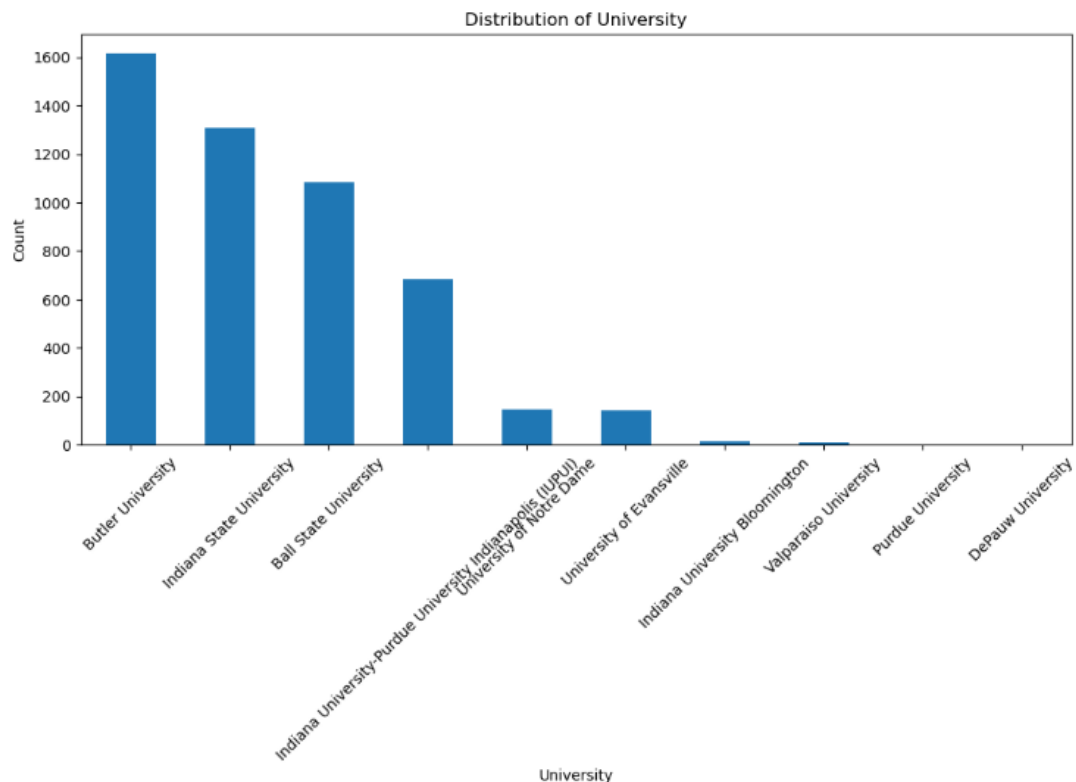
- Using Pandas to model the distribution of the University Varaible we are able to see the overwhelming amount of students from Butler, Indiana State, Ball State, and IUPUI universities.
    - This data could be used for allocation of resources and catering to a specific university, let's say for example, Butler which has more students than the next university by 300 students. More allocation of resources towards catering to butler university would mean an increase in sales, or satisfaction from nearly a third of the student population according to the data set.
- I also modeled the distribution of Orders, Majors, and Year. However I have no significant thoughts about how the data could be used for other business cases.

Consider implications of data collection, storage, and data biases you would consider relevant here considering Data Ethics, Business Outcomes, and Technical Implications

- **Discuss Ethical implications of these factors**
  - Privacy and Consent - We should ensure that students that whose data is collected gave informed consent for the use of their data.
  - Data Anonymization - We need to properly anonymize sensitive data to protect individuals privacy, and avoid collecting any identifiable information.
  - Data Security - Ensuring that the data is secure and is protected against unauthorized breaches, unauthorized access can lead to privacy violations.
  - Transparency - we need to maintain transparency about data collection and usage, people should know how their data is being used.

- **Discuss Business outcome implications of these factors**
  - Reputation and Trust: - Ethical data practices enhance trust. Data breaches or unethical data use can damage the organization's reputation and erode customer trust.
  - Legal and Regulatory Compliance - Violations of data privacy regulations can lead to legal and financial consequences.
  - Customer and Student Experience - Ethical data collection can enhance the customer or student experience. Using data for personalized services can improve satisfaction.
  - Bias Impact on Business Decisions - Biased data can result in poor business decisions. For example, if the dataset is biased, it can lead to biased recommendations, affecting user engagement and satisfaction.

- **Discuss Technical implications of these factors**
  - Data Storage and Encryption - Implement robust data storage and encryption practices to protect data. This includes secure storage, access control, and encryption of sensitive information.
  - Data Governance - Establish clear data governance practices, including data access controls, audit trails, and policies for data retention and disposal.
  - Algorithmic Fairness - Ensure that machine learning algorithms are designed to be fair and do not perpetuate biases. This may require algorithmic adjustments, which I don't think this work sample does.
  - Compliance with Data Privacy Laws - Technical systems must ensure compliance with data privacy laws, including mechanisms for data subject access requests and data deletion.

Given the work required to bring a solution like this to maturity and its performance, what considerations would you make to determine if this is a suitable course of action?

- Technical Feasibility: Assess the technical feasibility of the solution, including required technology, infrastructure, and expertise.
- Costs and Resources: Estimate the costs and resources required for the project to ensure financial viability.
- ROI and Business Impact: Calculate the potential return on investment (ROI) and assess the business impact.
- User Acceptance and Adoption: Engage with end-users to ensure that the solution aligns with their needs and preferences.
- Scalability and Integration: Ensure that the solution is scalable and can be integrated with existing systems and processes.
- Risk Assessment: Conduct a comprehensive risk assessment and identify potential risks and mitigation strategies.
- Pilot Testing: Consider conducting a pilot test or proof of concept before full-scale implementation.
- Feedback and Continuous Improvement: Establish mechanisms for collecting feedback from users and continuously improving the solution.