Project Topic: Job market in the United States in 2023

MIS 584

Business Intelligence Project:

JOB MARKET INTELLIGENCE FOR WPI HEEBNER CAREER DEVELOPMENT CENTER (CDC)

Authors

Nicolle Fasce

Yuan Feng

Baihua Huang

Aswin Kumar Bisati

Professor

Professor Kordzadeh

Fall semester 2023

Table of Contents

Project Topic: Job market in the United States in 2023

Table of Figures:

Project Topic: Job market in the United States in 2023

Section 1: Executive Summary

Data has always been a key driver to success in any field where fast and accurate decisions have to be made. It can represent a competitive advantage in any type of business, even more when data analytics became part of the business culture and motivates different areas to share and work together in order to obtain the best insights to improve performance (Ransbotham & Kiron, 2017). The main goal of this project "Job Market Intelligence" is to demonstrate to WPI Heebner Career Development Center (CDC), a WPI department located at Unity Hall - 5th Floor, 100 Institute Road in Worcester MA, that data analysis along with business intelligence implementation can guide and lead its efforts to better results.

WPI Heebner Career Development Center, better known as CDC, can be considered as a strategic department of WPI due that its main purpose is to organize events, workshops, fairs, conferences, and training for WPI's students in order to support their career development. This project is looking to offer a business intelligence solution by demonstrating how data analytics and visualization can generate valuable insights about the behavior of the job market in the United States in 2023, in order to support the customization of their services to what the labor market is expecting from WPI's students.

To achieve this goal, it has been used a "LinkedIn Job Postings – 2023" dataset generated in august 2023 by Arsh Kon, shared via Kaggle (Kaggle, 2023) to analyze company engagement and job-related aspects, such as skills, benefits, work types, experience levels, etc. This dataset offers 11 tables[1], including JOB_POSTINGS, SALARIES, BENEFITS, COMPANIES, EMPLOYEE_COUNTS, SKILLS, JOB_SKILLS, INDUSTRIES, JOB_INDUSTRIES, COMPANY_SPECIALITIES, and COMPANY_INDUSTRIES.

Finally, as part of this project it has been created an extra visualization tool based on WPI´s internal data "Student´s First Destination", with the objective of providing an example of how they can measure the impact of the implementation of business intelligence to structure their services.

Problem Statement

Limited Job Market Information:

CDC current difficulties in accessing job market data and real-time information about industry trends and demands means that they do not have a full picture of what is happening in today's business market. Furthermore, the data access and processing limitations constrain the ability to learn more about skill needs and salary trends across different jobs and positions. These restrictions also limit the ability to accurately predict future business market needs making it difficult to provide highly personalized career advice to WPI´s students based on their skills and aspirations. Working on expanding CDC resources can help students to navigate the ever-changing job market while offering more specific and structured services.

Define Accessibility Challenges

Data governance mandates, critical for compliance, may unintentionally limit access to valuable data, hampering informed decision-making. The necessity for robust security measures due to cybersecurity threats and privacy regulations can complicate data access and utilization. Diverse data formats and structures pose challenges for integration and analysis, demanding additional effort and resources for standardization. Limited data literacy among users hinders effective data access and utilization for meaningful insights. Budgetary constraints make implementing robust data accessibility solutions expensive, necessitating strategic resource allocation. Technological

[1] See Appendix 1 for external dataset descriptions.

limitations in data storage, processing power, and accessibility tools can impede effective data management and sharing across the organization.

How BI Can Help

Comprehensive Job Market Insights:

Implement a BI system to aggregate and analyse real-time job market data. Provide actionable insights into industry trends, skill demands, and employment opportunities.

Implementing a Business Intelligence (BI) system is crucial for overcoming limited business data challenges. It aggregates and analyses real-time job market data, providing insights into industry trends, skill demands, and employment opportunities. BI tools uncover hidden patterns and trends in job postings, salary information, and skill requirements, helping organizations identify skills gaps and plan workflows effectively. Additionally, BI facilitates the creation of customized career development strategies, enhancing employee engagement. By using BI in decision-making, organizations can make data-driven choices in talent acquisition, employee development, and human resource management, reducing risks and contributing to organizational success. Ultimately, BI implementation offers a comprehensive understanding of the job market, enabling informed decision-making and the development of a future-proof workforce for a competitive advantage in attracting and retaining top talent.

Efficient Data Utilization:

There are several key steps to creating a structured data environment using BI tools to make effective decisions in business development processes. To begin with, it is important to clearly define the objectives of performance development strategies and identify key parameters to be measured, such as employee performance and skill development. The next step is to integrate data from multiple sources, including HR systems and training systems, to ensure data quality and consistency across the implementation Choosing the right BI tools, such as Tableau or not Power BI, is important, considering things like scalability and compatibility with existing systems. Automation of data analytics tasks such as cleaning and collection is key, allowing HR professionals to focus on strategic activities Real-time visualization capabilities so enables organizations to react quickly to changing circumstances and make informed decisions. The use of BI tools to visualize data using intuitive dashboards makes it easier for stakeholders to understand complex data. Based on historical data, predictive analytics models help predict future trends in business development and proactively address potential challenges. User training ensures that relevant stakeholders can analyse and interpret data effectively. Prioritizing data security and regulatory compliance is essential and implementing features such as access control and encryption. Finally, a commitment to iterative improvement requires ongoing monitoring, data collection, and modifications to increase the quality of data sets and their relevance to business improvement strategies.

Data Warehouse Implementation:

Establish a centralized data warehouse for organized storage. Ensure data consistency and accessibility across the Career Development Centre. Business intelligence (BI) plays a key role in implementing data warehouses by making the system more flexible and efficient. Data modelling and design: BI tools provide comprehensive data modelling functionality and enable organizations to create data warehouses that efficiently store large amounts of data and organize and use them. Data Quality Management: BI tools provide data quality management features that help ensure accuracy, integrity, and consistency of data in the data warehouse

Project Topic: Job market in the United States in 2023

Goals

Optimizing Resource Allocation:

Utilize BI to better understand job market requirements. Align resources, programs, and support services based on identified trends and demands.

Establishing a centralized data warehouse for organized storage is a fundamental step in ensuring data consistency and accessibility across the Career Development Center. Business Intelligence (BI) plays a pivotal role in this process by enhancing the flexibility and efficiency of the system. BI tools provide comprehensive data modelling functionality, enabling organizations to create data warehouses capable of efficiently storing large volumes of data and organizing it for effective use. Additionally, BI tools offer features for data quality management, ensuring the accuracy, integrity, and consistency of data within the data warehouse. This comprehensive approach to data management, facilitated by BI, not only streamlines storage but also enhances the reliability and utility of the information stored within the centralized data warehouse.

Strategic Decision-Making

The primary objective of the Business Improvement Center is to achieve strategic decision-making through the use of business intelligence (BI) insights. By leveraging the power of BI tools, we aim to make informed decisions that improve the overall effectiveness of our business development process. With a data-driven approach, we intend to examine key metrics related to employee performance, skills development, and market growth. The use of BI tools will automate data analysis tasks, ensure efficiency, and free up resources for more strategic efforts. Real-time insights into operational trends and employee performance will allow us to quickly adapt to changing circumstances and optimize our operations for efficiency. Using complex data visualization techniques will translate complex data into easy-to-understand methods to help stakeholders make informed choices. Moreover, the integration of predictive analytics will enable us to actively address challenges, exploit emerging opportunities and strategically plan for future growth. Ongoing research, user training, and iterative improvement will be necessary to refine our data-driven strategies and achieve our goal of making impactful decision-making in career development.

Project Deliverables

Prototypes

- **Dashboard 1: For students**
    - Cluster the average normalized salary into 3 categories (high, medium, low);
    - Then we perform graphics for skills, industries, locations, experience requirements, and work types;
    - This dashboard aims to provide WPI graduates a comprehensive view of job market, helping with figuring out where they want their career to begin with.

- **Dashboard 2: For CDC Staff**
    - Categorize the industries into 3 sectors aligned with external standards
    - Then we perform graphics for job postings, company postings, and average salaries by industries and locations.
    - Further creating charts for top company engagements, top popularities, top preferred skills, and a list of specialties.
    - The second dashboard objective is to provide insights on which companies to invite, and identify the industry needs which helps in reduce the education-industry gap.

- **Dashboard 3: For CDC Staff**
    - Using operational or public data to create indicators like total graduates, the number of graduates employed, success rate, and so on;
    - The third dashboard aims to provide CDC staff an evaluating visualization on the graduates' outcomes.

The collective strength of these three dashboards lies in their capability to inform CDC's strategic planning and enhance career support for our students, thereby advancing their professional trajectories based on real-world market insights.

Conjoint Analysis

- Utilize conjoint analysis to understand the importance of several factors in job market trends, such as job location, company size, and industry sectors.
- Provide actionable recommendations based on the analysis.

Section 2: Introduction

Client Company background

Mission Statement

"The WPI Career Development Center's mission is to empower all undergraduates, graduate students and alumni to identify and achieve their career goals" (Career Development Center, 2023)

About WPI Heebner Career Development

The WPI Heebner Career Development Center (HCDC) is a vital entity within the realm of education and non-profit industries, dedicated to shaping the professional trajectories of WPI's students and alumni, as a flat organization with a lean team of 8 members, HCDC focuses on personalized information pushing, skill training, and orchestrating job affairs to enhance the employability of its constituents. Serving as an integral department at WPI, the HCDC operates within the expansive geographic market area encompassing all WPI students and alumni, striving to bridge the gap between academia and industry. The centre's commitment to fostering career growth and development underscores its significance within the broader educational landscape.

Products and Services of WPI Heebner Career Development

- Enterprise: CDC Employer Recruiting Services
    - Job Postings: Recruit WPI talent for a range of open positions by posting on CDC online platform.
    - Career Fairs/Career Expos: Participate in CDC annual Fall, Spring, and virtual Career Fairs or a career expo focused on a specific academic discipline.
    - Host a Company Event: Provide general information about business organization by coming to campus to host an event.
    - On-Campus Interviews: Interview job posting applicants by utilizing the interview suite at the Career Development Center.
    - Networking Events: Spend an evening interacting with students, alumni, and other professionals to exchange information, share advice, and develop contacts.
    - Educational Programs: Participate in a range of educational programs for students such as mock interviews or elevator pitch critiques.

    - Career-related Presentations and Panels: Participate in a panel or present on topics such as interviewing skills, how to work a career fair, the summer internship job search, and others.
    - Company Tours: As a great way to introduce students to business organization, the CDC will work with companies, faculty, and student organizations to arrange a tour.
    - CDC Staff Consultation: Meet with a CDC staff member to discuss a partnership strategy for business organization.
    - Office and Event Sponsorships: For maximum visibility to WPI students, CDC accepts office sponsorship.
- Student and Alumni: CDC specifically student support
    - Expanding students' knowledge of the broad range of career and employer possibilities.
    - Making connections with employers and alumni to understand the employment landscape, hiring process, and leverage students' network.
    - Recognizing international students have additional concerns such as work-authorization and visa sponsorship. CDC works in partnership with International House to guide students through the internship, co-op, and job search process.

CDC can provide advice to help students explore career options, write or update students' resumes, prepare for an interview, or negotiate a salary, the Heebner CDC is students "go to" collaborator and coach for career success. Get started by exploring the resources on the site.

- Search for jobs, internships, or co-ops in Handshake;
- Start the approval process for students' co-op or internship using the 'Experiences" tab in Handshake;
- View date, times, and employers attending upcoming WPI Career Fairs in Handshake;
- Attend drop-ins (M-F 12 pm – 4 pm at the HCDC) when classes are in session (no drop-ins in the summer) for quick questions or an initial resume review.
- Attend the HCDC Workshop Wednesday, and much more. (Ways to recruit, 2023)

Current BI status of WPI Heebner Career Development

The WPI Heebner Career Development Center (CDC) currently relies primarily on Microsoft Excel for its data management and analytics needs. Although Excel provides basic functionality, its limitations become apparent when dealing with large and complex data sets, preventing CDC from gaining the detailed perspective needed to support business improvement initiatives A dedicated data warehouse its absence no longer prevents the center from maintaining accurate data. Despite the use of Workday as an integrated information system, the lack of specific integrated information systems and a dedicated BI team, provided by a staff member named Ricky, is responsible for all BI responsibilities concerns arise about the effectiveness and feasibility of the CDC BI effort on the tool front, CDC uses Tableau and Excel to create data visualizations and to integrate data sources to meet reporting and visualization needs. While these tools enhance visualization capabilities, the sole reliance on administrators highlights the growth and development opportunities that may exist in establishing a more robust BI system at CDC. Aligning with industry best practices for effective business intelligence and adopting a collaborative, scalable approach to data-driven decision-making can further enhance CDC BI capabilities and help improve overall business growth in various systems.

BI Maturity Model of WPI Heebner Career Development

The WPI Heebner Career Development Center is currently in the prenatal to infant phase of a business intelligence (BI) growth model. The Center relies on performance reporting, and works

primarily with spreadmarts, which means a decentralized and manual approach to data processing. Reports can be generated on request, but the lack of automatic updates indicates potential challenges in obtaining real-time data. Going forward, there is an opportunity to grow and improve BI capabilities, such as implementing automated and integrated reporting systems, reducing reliance on manual data processing, enhancement of real-time data capabilities and progression to higher BI maturity stages will.

Related Case Studies

Case 1: Northern Virginia Community College - Labor Market Intelligence

In our pursuit of excellence at WPI's Career Development Center, we look to leading examples of field in job market intelligence that align with our goals at WPI's CDC. One of a related case study is Northern Virginia Community College's (NOVA) Labor Market Intelligence (LMI). NOVA's LMI exemplifies the strategic use of economic and demographic data to support their college's mission, which is preparing the workforce that meets the economic demands of the region (Northern Virginia Community College, 2023).

Their intelligence serves as a hub for various stakeholders, including faculty, students, local employers, and community organizations, offering access to up-to-date data products. With economic and demographic data, they have conducted in-depth research into economic conditions and employment trends, supporting the creation of a workforce that stands out on a global scale. Their outputs, including interactive dashboards and detailed reports, and provide valuable insights into in-demand sectors, hiring patterns, and critical economic indicators that all pertinent to NOVA's degree and continuing education programs. These insights hint on the strategies of curriculum development, career services, and community engagement.

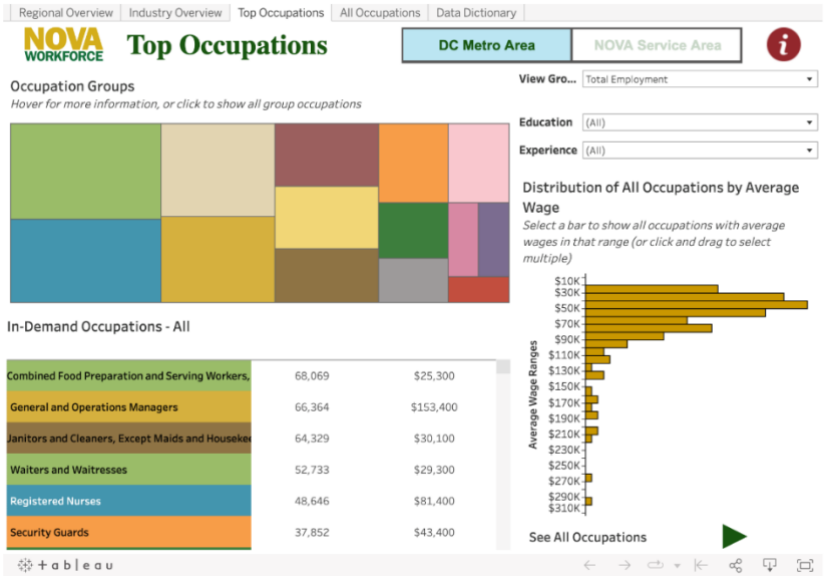


Figure 1 NOVA Labor Market Intelligence - Top Occupations

Project Topic: Job market in the United States in 2023

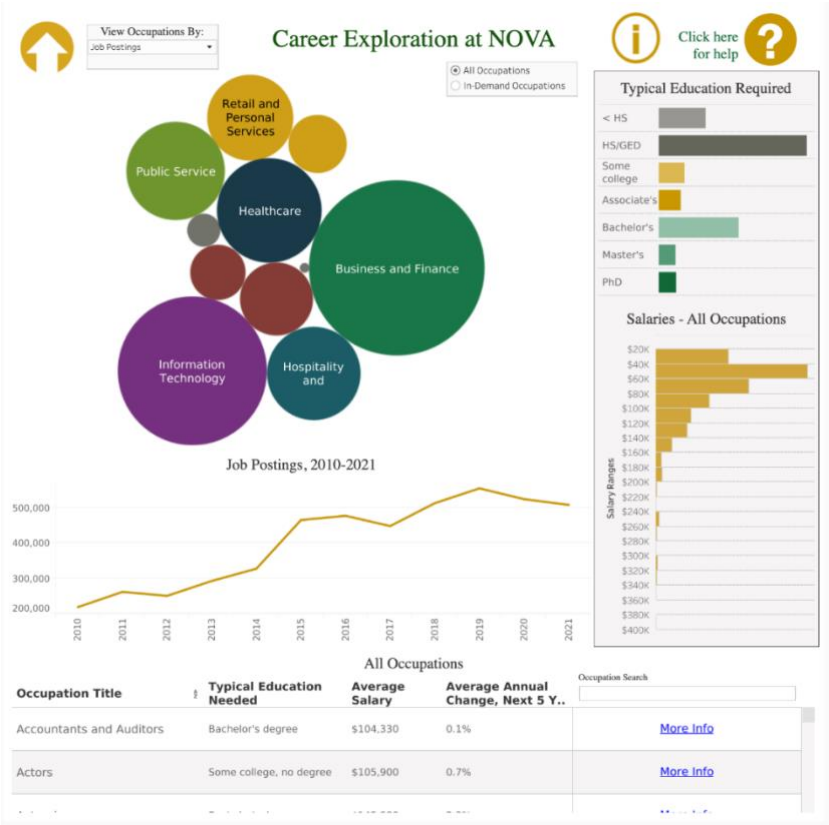


Figure 2 NOVA Labor Market Intelligence - Career Exploration

With similar mission, NOVA's LMI endeavours provide closer looks and great guidelines to our project about how we display the employment trends on WPI's CDC intelligence.

Case 2: Asobanca Financial Dashboard

Asobanca is a non-profit, private Ecuadorian organization whose main objective is to contribute to improving the quality of life for the entire Ecuadorian population. It works to strengthen the financial system by encouraging the implementation of best practices and compliance with all national and international guidelines. (Asobanca, 2023)

This NGO has developed a visualization tool called DataLab. The objective of this tool is to generate free and public information about the performance of the Ecuadorian financial system by sharing useful data to guide its users to decide how to allocate their resources into the financial system (Asobanca, 2022). All banks (Big, medium, and small) share every month their interest rates, financial statements, and other relevant indicators to feed this tool.

This BI implementation was a guideline for this project, not only proving the power of information in decision making, especially in the finance field, but also showing that data visualization is important and has to be easy to understand to all types of users.

Project Topic: Job market in the United States in 2023

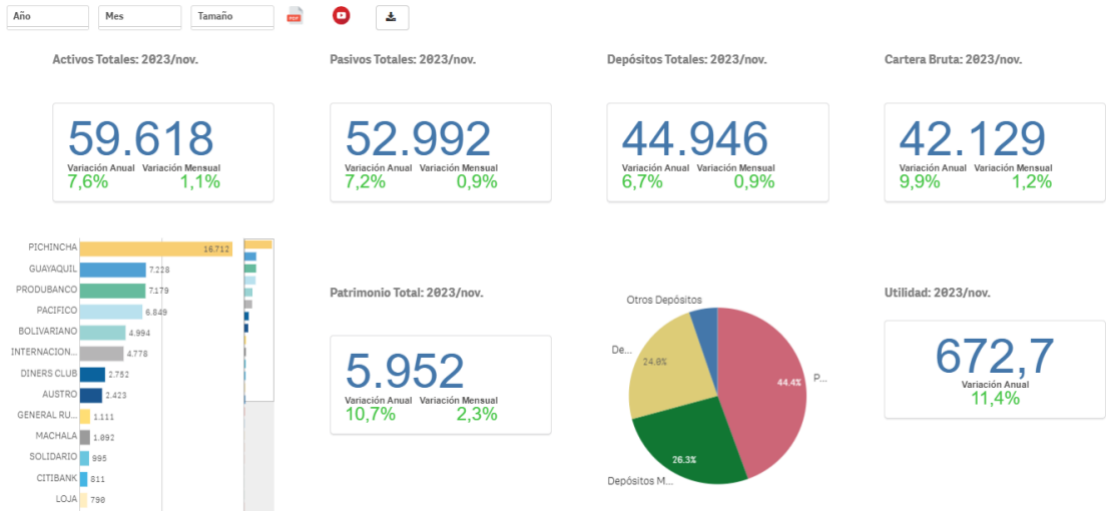


Figure 3 Banks performance information

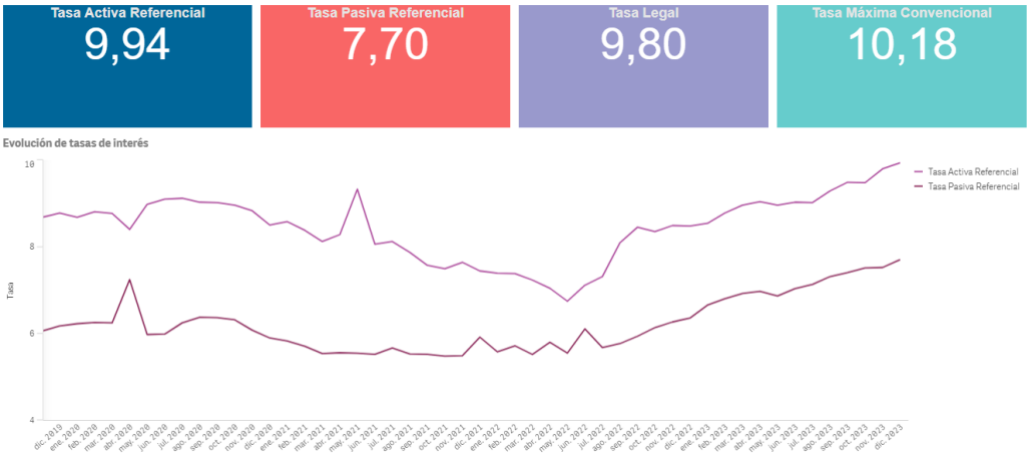


Figure 4 Interest Rates Evolution

Section 3: The Proposed BI Solution

What is Business Intelligence?

"Business intelligence (BI) is a term that describes a comprehensive, cohesive, and Integrated set of tools and processes used to capture, collect, integrate, store, and analyse data with the purpose of generating and presenting information to support business decision making." (Coronel & Morris, 2019)

What are the main components of BI solutions? What is role and value of each component?

The main components of any BI solution are based in the Business intelligence framework. This framework is a guideline not only to structure but to implement Business intelligence in any kind of business.

Project Topic: Job market in the United States in 2023

The following key components are detailed based on CDC framework necessity.

1. **Data Sources:** All data available that can be extracted and can be useful for the end of the BI solution.
    a. **Internal:** Workday, Handshake, etc.
    b. **External:** Social media channels like LinkedIn, Glassdoor, etc.
2. **ETL Process / Data Integration:** Transforming data into a unified format to be able to be analysed. In this stage it has to be considered data quality and integrity validation in order to move forward to the next step.
3. **Data Warehousing/Data Mart:**
    a. Centralized repository that stores consolidated and historical data from different sources. It is the foundation for reporting and analysis.
    b. CDC Star Schema to store structured data.
4. **Query and reporting:** Process of retrieving data from the data warehouse in order to report performance.
5. **Monitoring, Alerting, and Data Analysis:** process of data analysis to create valuable information and insights about the operations and performance of any business.
6. **Visualization:** process of visualizing processed data throughout visualization tools sharing the information in a friendly and easy way to understand.

Business areas involved in the BI Framework:

i. **People:** Analysts, other teams (IT), users of any business influence in data transformation, interpretation, and other aspects of BI implementation.
ii. **Management:** team leaders, managers, directors influence in approvals and decision making of the BI implementation.
iii. **Governance:** standards, and regulations for the privacy and control of the data available.
iv. **Processes**: includes the collection, cleaning, structuring, integration, and transformation of data to be available for query and reporting.

All these components collectively contribute to create a robust BI environment that can empower organization to turn raw data into actionable insights for strategic decision making (Eckerson, 2010), and all of them are applicable to CDC.

Project Topic: Job market in the United States in 2023

CDC Framework components:

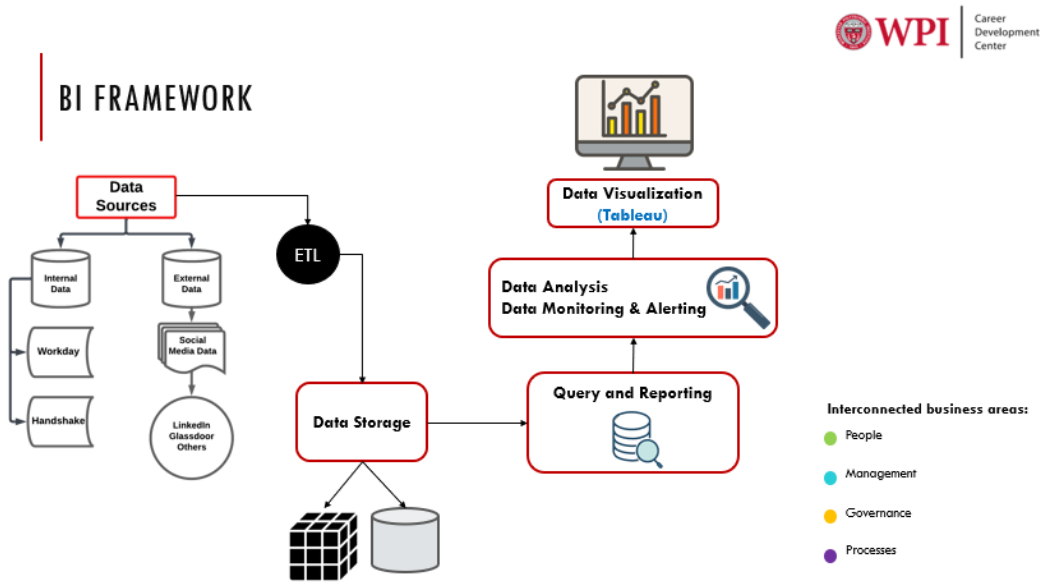


Figure 5 BI Framework applicable for CDC

Data warehouse Schema:

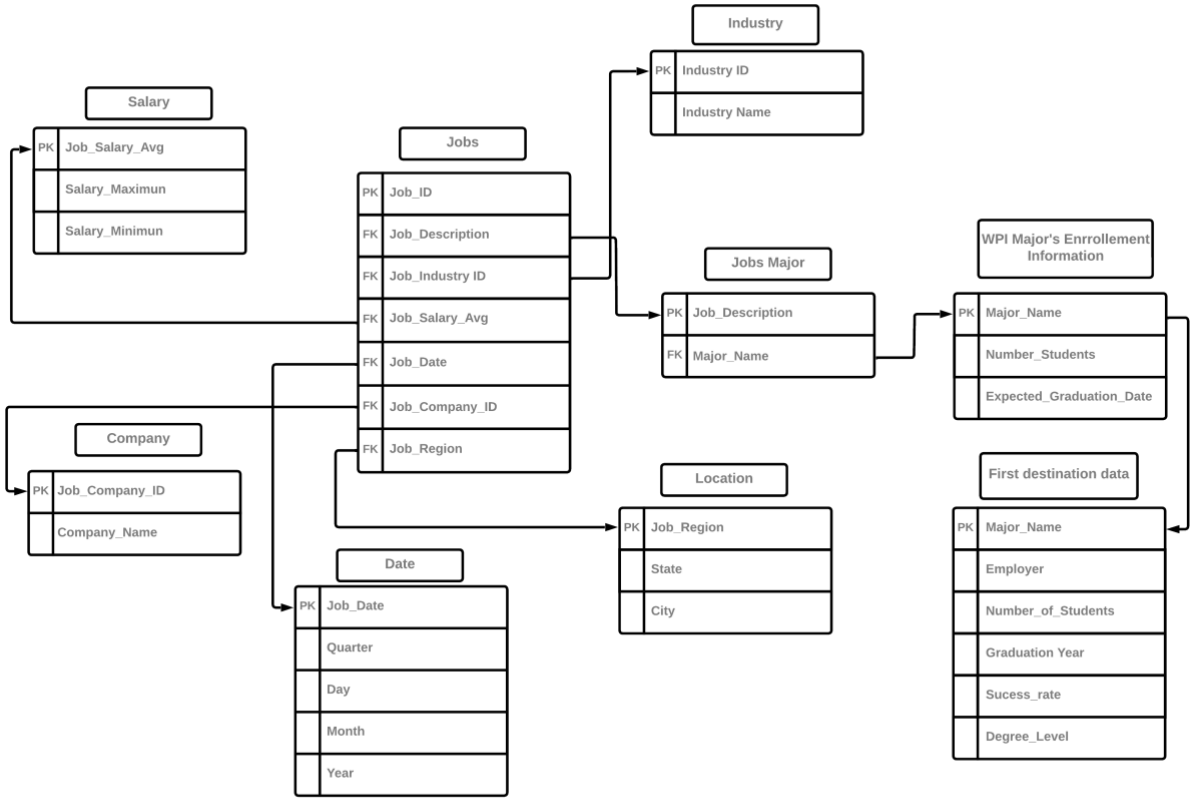


Figure 6 Data warehouse Star Schema applicable for CDC

Section 4: Dashboards Prototypes

Dashboard 1: Where do you want your career to begin?

The first dashboard serves as tactical instrument for WPI's students. This tool answers the first question all students ask themselves when are looking for their first job. Where do you want your professional career to begin? This dashboard is an example of how useful BI can be, and how visualization is important to present relevant information for decision making. This dashboard is the kind of dashboard that CDC could develop for WPI´s students if they decide to implement business intelligence to support their services. The information available in this dashboard helps students to navigate through LinkedIn job postings in a much easier way than going to LinkedIn and try to find the job they would like. This tool is a summarized dashboard that can guide you through out each of the United Stated job's locations, identifying the average salary and number of jobs posted by state, by industry, by experience level, and helping you to discover the type of job (Full time, part time, contracts) available, showing you the companies that are posting those jobs, also identifying which benefits are offered in that particular job and the skills required by the companies and industries.

This tactical dashboard prototype could become the main tool for job searching for all WPI´s student, saving time and efforts, exploring all options available centralized in one single screen.

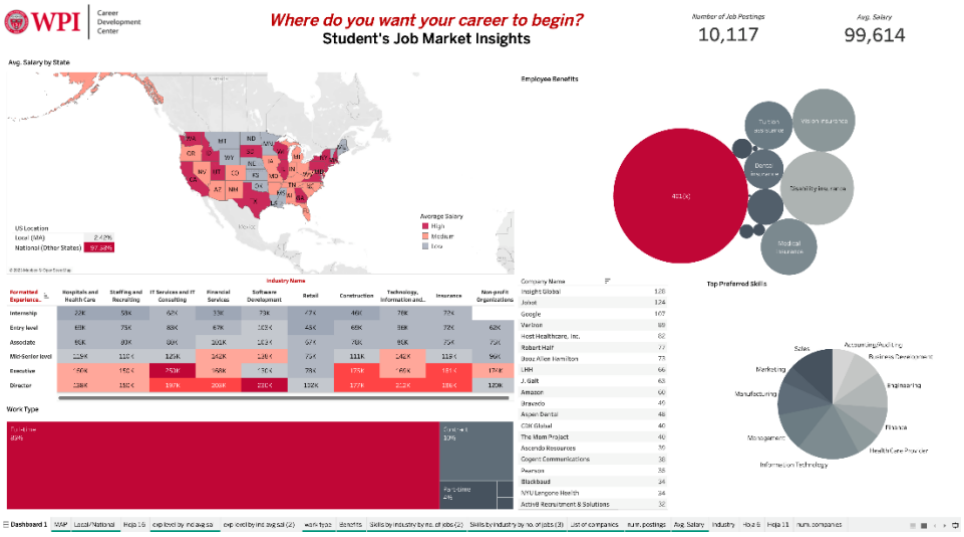


Figure 7 Dashboard 1: Where do you want your career to begin?

Dashboard 2: Company Engagement & Strategic Insights from Job Market

The secondary dashboard serves as a tactical instrument for the WPI's CDC staff to identify ideal participants for career fairs and discerning company hiring trends. It concentrates on three primary metrices: the number of active hiring companies (Company Postings), the volume of job vacancies (Job Postings), and the prospective compensation (Average Salary). On the dashboard, these metrics are summarized at the top left, complete with a 'Select Data' filter for choosing which metric to view. Additionally, a 'Select Category' filter allows for more specific industry analysis, with options ranging

from a comprehensive view ('All') to focused sectors like 'Arts, Humanities, Social Sciences and Interdisciplinary', 'Business, Finance, and Legal Services', and 'Technical, Engineering, and Science'.

On the dashboard's left, we have a bar chart breaking down the selected metric and sector by industry, following with a map that showcases the geographic distribution by tailored data. To the right, four distinct metrices are visualized. The first two charts: a bar chart listing the top engagement companies, a rectangular matrix of the most popular companies. Integration with further analysis, such as conjoint analysis, if possible, might enriching our strategies for inviting companies that align with our students' preferences, ultimately aiming to enhance the success rate of our graduates in securing their first roles. The other two charts - a bubble chart for the most in-demand skills by companies and a table detailing the specialties of these tailored companies), are critical for shaping training sessions to ensure our students are well-prepared to enter the job market.

For practical application, say we aim to expand job opportunities in technical fields. We would apply the ‘Job Postings’ and ‘Technical, Engineering, and Science’ in filters. This action tailors the displayed data to relevant industries, for example, that the ‘Technology, Information and Media’ industry offers 1345 jobs, while ‘Manufacturing’ has 1035 jobs, and so forth. Moreover, the map displayed the number of job postings in each state, for example, Massachusetts provides 75 related jobs. If we are keen on engaging more local enterprises, selecting MA on the map will further reveal the local opportunities by the four metrices.

For example, we can easily identify the highest talents in-demand companies including Google provides 107 jobs, Amazon provides 60 jobs, Bravado provides 49 jobs, etc. Their history of participation in our fairs would show on the tooltips (if integrate with the career fair operational data). Additionally, the dashboard visualizes which companies are most followed as well as their participation history, for instance, Google with 135 million followers, Amazon with 129 million followers, Tesla with 59 million followers, and so forth. On the other hand, it displays the most sought-after skills based on the tailored data, like 279 companies required ‘Information Technology’ related skills, 230 companies required ‘Sales’ related skills, etc. Moreover, the specialties chart reveals the areas in which these companies specialized, such as, Google specialize in ads, android, apps, artificial intelligence, etc. These two charts help with better designing training sessions that effectively prepare our students for successful career starts.

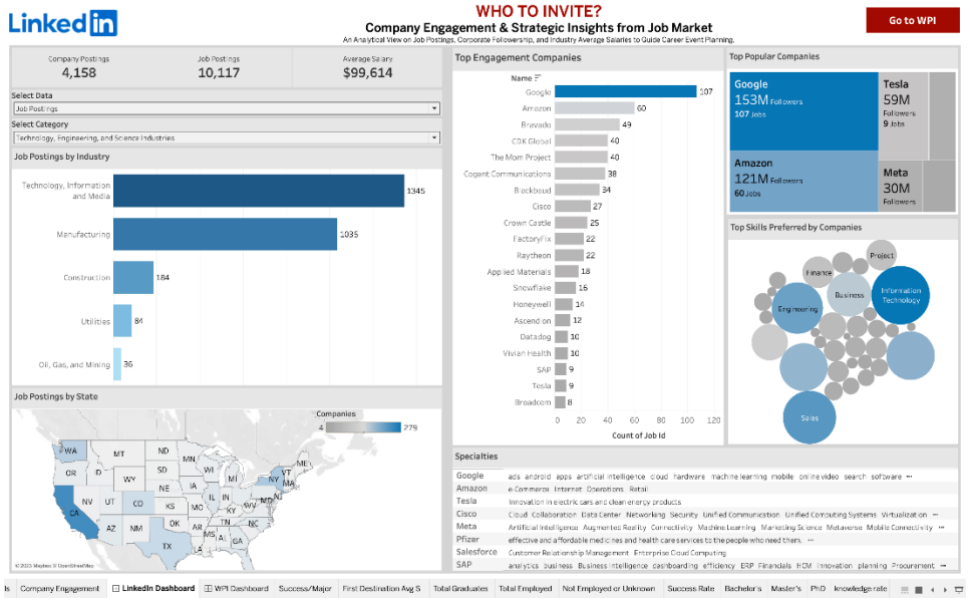


Figure 8 Dashboard 2: Company Engagement & Strategic Insights from Job Market

Project Topic: Job market in the United States in 2023

Dashboard 3: First Destination Outcomes

Our third dashboard presents an in-depth analysis of WPI graduates' first destination outcomes, offering a historical data set that can be filtered by graduating class. This dynamic dashboard includes several key performance indicators: the total number of graduates, the active majors count, employment and further study statistics, service participation, the number seeking employment versus those not, and the success and knowledge rates. It also provides information on average starting salaries and the diversity of employers.

Through this dashboard, we are equipped to address queries about our graduates' performance in the job market. For instance, by applying the same sector as the second dashboard, which is the 'Technical, Engineering, and Science' sector, we can observe the distribution of job placements across different degree levels through pie charts, compare starting salary, review success rate by various majors and degree levels. Additionally, it offers a detailed list of employers engaged with our graduates.

Moreover, this tool is instrumental in highlighting areas for enhancement. If a particular major exhibit underperformance, we can delve into the specifics by clicking it in the bar chart of success rate by major, to devise strategies aimed at better supporting students in those disciplines.

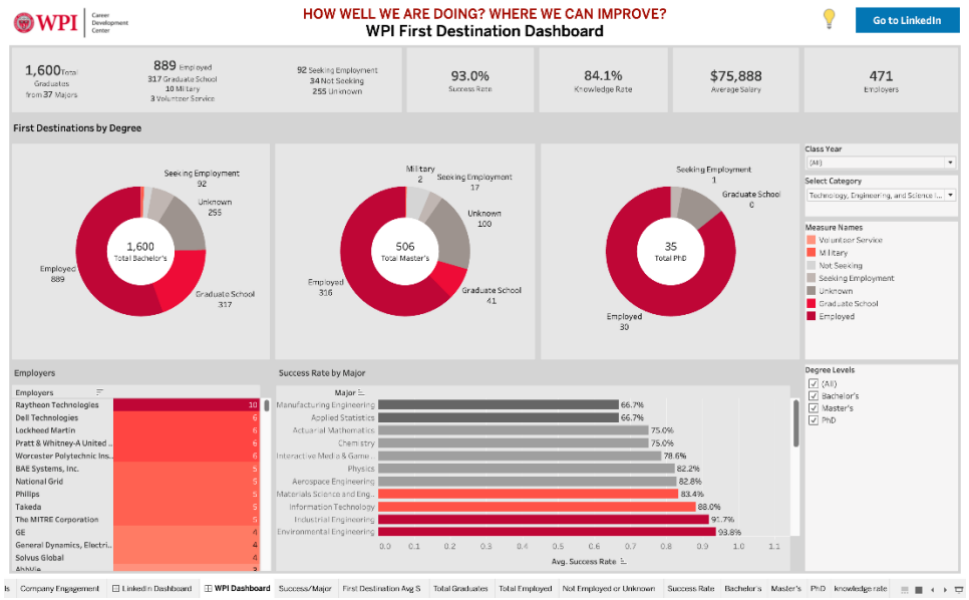


Figure 9 Dashboard 3: First Destination Outcomes

Conjoint Analysis: Identifying students' job´s preferences

In the Conjoint Analysis dashboard, we dive into a thorough examination involving 34 students from Worcester Polytechnic Institute (WPI). To gather insights, we used a questionnaire with 12 questions and a 1 to 10 rating scale. Our goal was to figure out what influences these students' decisions the most when it comes to academic and career choices. We are looking at three main things: where they want to be (Local or National), the size of the companies they are interested in (Small to Medium or Large), and the industries they are considering (like Technology, Engineering, Science; Business, Finance, Legal Services; and Arts, Humanities, Social Sciences, Interdisciplinary Studies). It was used Excel and Tableau to help us make sense of all this data. See appendix 4 about the conjoint

analysis survey and the regression results. What we found is that students are particularly drawn to the Technology, Engineering, and Science Industries, scoring the highest at 1.925. Also, the overall industry category holds the most weight, with a significant 63.38% in terms of importance. This will help CDC in their decision making and better assist students affairs.

Factors Coefficient

Factor	
Local	0,185
National	0,000
Small to Medium Company Size	-0,928
Large Enterprise/Corporation	0,000
Technology, Engineering, and Science Industries	1,925
Business, Finance, and Legal Services	1,171
Arts, Humanities, Social Sciences, and Interdisciplinary Studies	0,000

Figure 10 Conjoint Analysis Utilities

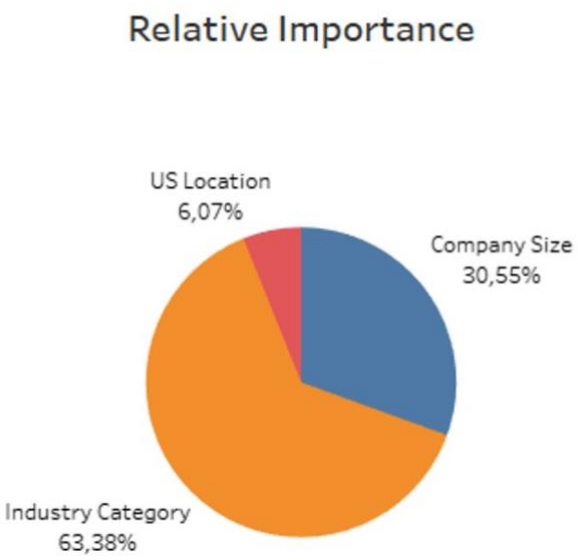


Figure 11 Conjoint Analysis Attributes' Importance

Section 5: Implementation

Managerial implementation

To explain how the BI implementation can be handle from the managerial perspective, for this project was analysed the Kotter's eight step model. The following table explains each step CDC should go through to implement BI as part of its structure.

Project Topic: Job market in the United States in 2023

Table 1 Kotter's Managerial Eight Steps

Kotter's eight steps	Career Development center (CDC) - BI implementation
Establish a sense of urgency	• 100% of CDC Data is not structured and is not easily available to analyze and visualize. • Difficulty understanding student's needs and job market requirements.
Form a powerful guiding coalition	• CDC Director, and the Assistant Director of Career Services can convince WPI directors to support a process of BI implementation by showing them the importance to understand how to better allocate their resources to increase the student's success in their first destination job.
Create a Vision	• BI implementation for CDC represents more than just information available to support their services. It is an opportunity to find new ways to collaborate with students to improve their first job experience. It is also possible to create a historical record to analyze the job market behavior and promote forecasting to create new strategies for future years' classes. Also, it will be possible to find new trends in the job market to adjust program's curriculums. Data sources, data warehousing, and real-time visualization are needed to support this and more related initiatives.
Communicate the Vision	• CDC needs to share a clear and consistent message about their initiative of BI implementation throughout all internal communications means of WPI, with the main objective of gaining supporters, and new users and beneficiaries of this new source of information.
Empower others to act on the vision	• Admissions, undergraduate and graduate student's offices can help and support the project to use this data to better guide students during their experience in WPI.
Plan for and create short-term wins	• Year over year evaluation of success rate in first destination job analysis. • Satisfaction surveys from students and CDC services.
Consolidate improvements and produce still more change	• Use CDC data warehousing for other types of analysis. For example, to analyze which are the best companies to work for WPI students based on student's preferences. Identify which companies should be invited to the career fair to improve recruitment.

Institutionalize the new approaches	• Increase data sources for CDC data warehouse to improve data analysis and expand the scope of the BI implementation.

(Watson, Goodhue, & Wilxom, 2002)

Technical – Data quality problems

The data quality challenges in this project mainly in three aspects: data completeness, data validity, and data consistency: (What is data quality? | IBM, n.d.)

Completeness

There are data completeness issues in both external and operational data. For external data in LinkedIn dataset, there are a lot of missing values, we did some basic data cleaning like:

- **Unrelated Variables:** Drop variables that are not pertinent to our project
- **Essential numerical variables** such as 'company_id', 'max_salary', and 'min_salary' will be closely examined. Rows with missing values in these fields will be dropped due to the lack of critical information.
- **Categorical variables** like 'formatted_experience_level', where missing values are significant in number, we have introduced a "Not Specified" category, which preserved the dataset's integrity without losing substantial information.
- See appendix 2 for data cleaning code.

For operational data, we fail to obtain any data, such as collected data during previous career fairs and training sessions, as well as employment data. Instead, we use the latest public data from The National Association of Colleges and Employers (NACE) about WPI's first destination report on class of 2020 to share CDC how to track their outcomes.

Validity

There are variations in scales for several columns in LinkedIn dataset.

- **Standardize the Salaries:** Given the variation in salary pay periods (hourly, weekly, monthly, yearly), we have established a 'normalized salary' column. This standardizes salary data, enabling accurate average salary calculations.
- **Standardize the Job Locations:** The job locations are recorded discrepancy that some of them with detailed address, others not. We cleared and separated the job locations into State and City, ensuring the clarity and usability.

Consistency

The data consistency issues occurred when we try to connect external data and operational data.

- **Unified Categorization for Industries and Majors:** Since the current obtained datasets do not have any field connects industries and degrees/programs at WPI, and we aim to present a comprehensive view that meaningfully connects both. To this end, we have developed a unified category field applicable to industries and majors. But there it is the challenge of data consistency.
- **Alignment with External Standards:** We have aligned our categorization with the external standards such as Industry Codes V2 (AmanTechWriter, 2023) and the Schools, Departments & Programs list (Schools, Departments & Programs | Worcester Polytechnic Institute, n.d.) to

ensure our categories are sufficiently broad enough to encompass a diverse range of fields, yet specific enough to offer valuable insights.

- **Handling Overlapping Categories:** We recognize that some industries and majors naturally span multiple categories. Currently, each industry is assigned to a single category, but it is important to continuously revisiting and refining this categorization. This iterative process will ensure our approach remains relevant, thoughtfully designed, and effective, especially in accommodating emerging and interdisciplinary fields.

Ethical – Privacy, and consequences of use of data

Privacy Concerns

Currently, we do not have privacy concerns for this project as all the data are from public. However, following aspects of ethical challenges that CDC should consider when they utilize the intelligence with up-to-date data:

- Comply with social/network media ethical and privacy policies when obtaining job market data from external sources.
- Comply with privacy laws as well as privacy policies at WPI, including ask for consent from users, inform the properly collect and use of the data, etc. when personal or sensitive data from WPI CDC operational processes are used.

Data Misinterpretation

There is a risk of misinterpreting data, especially in complex fields like the calculation in the number of job postings by industry, location, and companies, which can lead to misguided career advice or policy decisions. Understanding the source, industry coverage, and any source biases are crucial for CDC staff. For example, some industries may be over-represented (like tech and finance), and other may be under-represented (like arts) due to the nature of LinkedIn platform's user base and overlapping industries, which are subject to the selection bias during the data collection. (Bias in data analysis, n.d.) Also, we use job industries rather than company industries in the dashboard, therefore, it is normal that we see tech companies like Apple listed specialities in 'Arts, Humanities, Social sciences, and interdisciplinary' sector. On the other hand, acknowledging that LinkedIn's user base may not be a perfect representation of the overall job market. It might skew towards certain professions, educational backgrounds, or demographic groups. Therefore, CDC should consider using multiple sources for obtaining job market information if possible and think about the outcomes carefully before making interpretations and providing any career advice.

Maintaining Data Integrity

Besides regularly revisiting, cleaning and standardizing data to avoid biases or distortions, there are still challenges that CDC staff would face in maintaining its integrity during the utility of the intelligence. Since job data from platforms like LinkedIn is largely user-generated, which means the accuracy of the information depends on the users' honesty and attention to detail, and it is common to see different job postings compared with other sources filled by different hiring managers. Therefore, while the dataset provides valuable insights, it is crucial to interpret the data considering these factors to avoid bias in interpretation and ensure a comprehensive understanding of the job market dynamics.

Section 6: Summary and Recommendations

Results:

Conjoint analysis proves invaluable in understanding students' preferences for job features, enabling Career Development Centers (CDC) to identify crucial attributes influencing choices. The insights derived inform product/service development, allowing CDC to design offerings aligning with student needs. On the other hand, Business Intelligence (BI) tools facilitate informed decision-making by providing access to timely data, aiding trend analysis, and offering performance monitoring in areas such as salary levels and job industries. Data visualization through tools like Tableau enhances comprehension across the job market. Efficient data access and management not only ensures data integrity, supporting reliable decision-making, but also addresses compliance and security concerns, vital for safeguarding student privacy and maintaining trust. Streamlined processes contribute to operational efficiency, enabling CDC teams to access necessary data promptly, reducing delays, and enhancing overall productivity for effective student support.

The Utility of Conjoint Analysis

Understanding Student Preferences:

Conjoint analysis can be used in market research to understand students' preferences for different job features. It helps CDC identify the most important attributes and levels that influence students' choices.

Product Development:

CDC can use the insights gained from conjoint analysis to inform product development strategies. By knowing what features students value the most, CDC can design products that better meet students' needs and preferences.

The Utility of BI

Informed Decision-Making

BI tools and systems can help CDC or students to take informed decisions by providing access to relevant and timely data. Decision-makers can analyze trends, patterns, and key performance indicators (KPIs) to guide their strategies.

Performance Monitoring

BI can support CDC in monitoring the performance of various aspects, such salary level, job industry, and skills level. Dashboards and reports enable quick assessments of key metrics, helping CDC and students identify areas for improvement.

Data Visualization:

BI tools often include powerful data visualization capabilities, here we use tableau to make it easier for CDC and students to comprehend complex data. Visual representations can enhance communication and understanding across different departments.

The Utility of Data Access and Management

Data Integrity:

Efficient data access and management ensure the integrity of data throughout its lifecycle. By directly involving accurate, consistent, and up to date data warehouse it will support reliable decision-making.

Compliance and Security:
Proper data access and management strategies help organizations comply with data protection regulations and ensure the security of sensitive information. This is crucial for protecting Student privacy and maintaining trust.

Operational Efficiency:
Streamlined data access and management processes contribute to operational efficiency. Teams can access the data they need when they need it, reducing delays and improving overall productivity.

By conducting this research, our aim is to assist the CDC in obtaining the following outcomes.

- Easy way to understand problems and how to save efforts and resources.
- Future implementation and improvements (Data warehousing – structure data for continuous analysis, to help them to increase their hiring rate of WPI students "First destination rate.")

Future Opportunities

- Personalized Career Guidance: Using data analytics to provide personalized career guidance to students, based on real-time labour market trends and individual skill sets.
- Curriculum Influence and Advocacy: Using insights from data to influence educational curriculum, advocating for changes that align education more closely with industry needs.
- Continuous Learning and Adaptation: The landscape of work and education is continuously evolving. Ongoing learning and adaptation of the data analytics will be important to staying relevant and effective.

References

AmanTechWriter. (2023, July 4). *Industry Codes v2 - LinkedIn*. Retrieved from Microsoft Learn: https://learn.microsoft.com/en-us/linkedin/shared/references/reference-tables/industry-codes-v2

Asobanca. (2022). *Sistema de información financiera*. Retrieved from DataLab: https://datalab.asobanca.org.ec/datalab/home.html

Asobanca. (2023). *Sobre Nosotros*. Retrieved from Asobanca: https://asobanca.org.ec/asobanca/

Bias in data analysis. (n.d.). Retrieved from Codecademy: https://www.codecademy.com/article/bias-in-data-analysis

Career Development Center. (2023). *About the CDC*. Retrieved from WPI Career Development Center: https://wp.wpi.edu/cdc/about/#:~:text=The%20WPI%20Career%20Development%20Center's,and%20achieve%20their%20career%20goals.

Coronel, C., & Morris, S. (2019). Business Intelligence and Data Warehouses. In C. Coronel, & S. Morris, *Database Systems: Design, Implementation, and Management* (p. 590). Boston: Cengage.

Eckerson, W. W. (2010). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business (Edition 2).* Wiley.

First Destination Outcomes Report: Data Reinforces Value of STEM Education in Career Success. (n.d.). Retrieved from WPI: https://www.wpi.edu/news/first-destination-outcomes-report-data-reinforces-value-stem-education-career-success

Northern Virginia Community College. (2023). *EXPLORE THE NOVA REGION*. Retrieved from Northern Virginia Community College: https://www.nvcc.edu/osi/labor-market/explore-nova-region.html

Ransbotham, S., & Kiron, D. (2017). Analytics as a Source of Business Innovatio. *MIT Sloan Management Review*.

Schools, Departments & Programs | Worcester Polytechnic Institute. (n.d.). Retrieved from https://www.wpi.edu/academics/schools-departments-programs

Students – Career Development Center Worcester Polytechnic Institute. (n.d.). Retrieved from Career Development Center Worcester Polytechnic Institute: https://cdc.wpi.edu/channels/student/

Watson, Goodhue, & Wilxom. (2002). Kotter´s model for organizational transformation. *Information & Management (39)*.

Ways to recruit. (2023, August 9). Retrieved from Career Development Center Worcester Polytechnic Institute: https://cdc.wpi.edu/ways-to-recruit/

What is data quality? | IBM. (n.d.). Retrieved from https://www.ibm.com/topics/data-quality

Project Topic: Job market in the United States in 2023

Appendix 1: EXTERNAL DATASET DESCRIPTION

LinkedIn Dataset

Table Name	Description
JOB_POSTINGS	Contains details about job postings, including job ID (primary key), company ID (foreign key), title, description, salary information, work type, location, application details, listing times, and other related attributes.
SALARIES	Stores information about salaries for various jobs, linked to the JOBS table by job ID (foreign key). It includes details like salary_id (primary key), maximum, median, and minimum salaries, pay periods, and compensation types.
BENEFITS	Details the benefits associated with each job. It specifies the type of benefit and whether it was explicitly mentioned or inferred: Type (primary key) Job_id (reference job_postings table and foreign key)
COMPANIES	Provides comprehensive information about companies, including company ID, name, description, size, location details, LinkedIn URL, and other relevant data. company_id (primary key)
EMPLOYEE_COUNTS	Includes data about the number of employees and followers for each company, linked by company ID. It also records the time when this data was collected. time_recorded (primary key) company_id (references companies table and primary key)
SKILLS	Contains a list of skills with abbreviations and full names, serving as a reference for skills required in various jobs: skill abbreviation (primary key) skill name
JOB_SKILLS	Links specific skills to job postings, referencing the JOBS and SKILLS tables. It pairs job IDs with skill abbreviations: job ID (references jobs table and primary key) skill abbreviation (references skills table)
INDUSTRIES	Lists various industries with unique industry IDs and names: industry ID (primary key) industry name
JOB_INDUSTRIES	Connects jobs to industries, with each record linking a job ID to an industry ID: job ID (references job_postings table and primary key) industry ID (references industries table)

Project Topic: Job market in the United States in 2023

COMPANY_SPECIALITIES	Records specialities for each company, linking company IDs to their respective specialities: company ID (references companies table and primary key) speciality ID
COMPANY_INDUSTRIES	Links companies with industries: company ID (primary key) industry ID (primary key)

JOB_POSTINGS_Cleaned

Column Name	Descriptions
job_id	Contains 10,119 of unique job id
company_id	Contains 4,159 of unique company id
title	Job title
max_salary	Maximum salary
med_salary	Median salary
min_salary	Minimum salary
pay_period	Pay period for salary (Hourly, Weekly, Monthly, Yearly)
formatted_work_type	Type of work (Fulltime, Parttime, Contract, Internship, Temporary, Other)
location	State, City
applies	Number of applications that have been submitted
formatted_experience_level	Job experience level (internship, entry, mid-senior, associate, director, executive)
normalized_salary	Average salary after standardization, all scaled in yearly payment, calculated by (max_salary + min_salary)/2
city	The city of the job located
state	The state of the job located

First Destination from Class of 2020 Dataset

https://www.wpi.edu/sites/default/files/inline-image/Offices/Career-Development-Center/CDC_FirstDestinationOutcomesReport_2020.pdf

Appendix 2: Data Cleaning Process

LinkedIn Job Postings - data cleaning

In this notebook I will attempt to cleaning the job_postings data.

Importing Dependencies

In [1]:
```
# pip install wordcloud
```

In [2]:
```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt # for plotting chart
from functools import reduce # module to help merge multiple dataframe
from wordcloud import WordCloud # module to print word cloud
```

Data Collecting & Pre-Processing

In [3]:
```
# main dataframe
job_postings_data = pd.read_csv("job_postings.csv")
job_postings_data
```

Out[3]:

	job_id	company_id	title	description	max_salary	med_salary	min_salary	pay_period	formatted_work_type	location	...	closed_time	formatted_experience_level	sk
0	3757940104	553718.0	Hearing Care Provider	Overview\n\nHearingLife is a national hearing ...	NaN	5250.00	NaN	MONTHLY	Full-time	Little River, SC	...	NaN	Entry level	
1	3757940025	2192142.0	Shipping & Receiving Associate 2nd shift (Beav...	Metalcraft of Mayville\nMetalcraft of Mayville...	NaN	NaN	NaN	NaN	Full-time	Beaver Dam, WI	...	NaN	NaN	
2	3757938019	474443.0	Manager, Engineering	\nThe TSUBAKI name is synonymous with excellen...	NaN	NaN	NaN	NaN	Full-time	Bessemer, AL	...	NaN	NaN	B I M En
3	3757938018	18213359.0	Cook	descriptionTitle\n\n Looking for a great oppor...	NaN	22.27	NaN	HOURLY	Full-time	Aliso Viejo, CA	...	NaN	Entry level	
4	3757937095	437225.0	Principal Cloud Security Architect (Remote)	Job Summary\nAt iHerb, we are on a mission to ...	275834.0	NaN	205956.0	YEARLY	Full-time	United States	...	NaN	Mid-Senior level	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
33241	133114754	77766802.0	Sales Manager	Are you a dynamic and creative marketing profe...	NaN	NaN	NaN	NaN	Full-time	Santa Clarita, CA	...	NaN	NaN	
33242	108965123	NaN	Office Administrative Assistant	A fast-fashion wholesaler, is looking for a fu...	NaN	NaN	NaN	NaN	Full-time	New York, NY	...	NaN	NaN	
33243	102339515	52132271.0	Franchise Owner	DuctVentz is a dryer and A/C – heat vent clean...	NaN	NaN	NaN	NaN	Full-time	Greater Boston	...	NaN	NaN	
33244	85008768	NaN	Licensed Insurance Agent	While many industries were hurt by the last fe...	52000.0	NaN	45760.0	YEARLY	Full-time	Chico, CA	...	NaN	NaN	
33245	3958427	630152.0	Stylist/ Clorist	Karen Marie is looking for an awesome experien...	80000.0	NaN	35000.0	YEARLY	Full-time	Chicago, IL	...	NaN	NaN	s a

33246 rows × 28 columns

let's check if there is duplicate and null in `job_id`

In [4]:
```
duplicates = job_postings_data['job_id'].duplicated()
num_duplicates = duplicates.sum()
num_duplicates
```

Out[4]:
```
0
```

In [5]:
```
job_postings_data.isnull().sum()
```

Out[5]:
```
job_id                        0
company_id                  654
title                         0
description                   1
max_salary                22135
med_salary                31005
min_salary                22135
pay_period                19894
formatted_work_type           0
location                      0
applies                   17008
original_listed_time          0
remote_allowed            28444
views                      7360
job_posting_url               0
application_url           12250
application_type              0
expiry                        0
closed_time               32074
formatted_experience_level 9181
skills_desc               32909
listed_time                   0
posting_domain            13558
sponsored                     0
work_type                     0
currency                  19894
compensation_type         19894
scraped                       0
dtype: int64
```

Drop useless columns for our project

In [6]:
```
chosen_columns = ['job_id', 'company_id', 'title', 'max_salary','med_salary', 'min_salary', 'pay_period','formatted_work_type',
                  'location', 'applies', 'formatted_experience_level', 'work_type']
new_data = job_postings_data[chosen_columns]
new_data
```

Project Topic: Job market in the United States in 2023

Out[6]:

	job_id	company_id	title	max_salary	med_salary	min_salary	pay_period	formatted_work_type	location	applies	formatted_experience_level	work_type
0	3757940104	553718.0	Hearing Care Provider	NaN	5250.00	NaN	MONTHLY	Full-time	Little River, SC	NaN	Entry level	FULL_TIME
1	3757940025	2192142.0	Shipping & Receiving Associate 2nd shift (Beav...	NaN	NaN	NaN	NaN	Full-time	Beaver Dam, WI	NaN	NaN	FULL_TIME
2	3757938019	474443.0	Manager, Engineering	NaN	NaN	NaN	NaN	Full-time	Bessemer, AL	NaN	NaN	FULL_TIME
3	3757938018	18213359.0	Cook	NaN	22.27	NaN	HOURLY	Full-time	Aliso Viejo, CA	NaN	Entry level	FULL_TIME
4	3757937095	437225.0	Principal Cloud Security Architect (Remote)	275834.0	NaN	205956.0	YEARLY	Full-time	United States	NaN	Mid-Senior level	FULL_TIME
...	...	...	...	...	...	...	...	...	...	...	...	...
33241	133114754	77766802.0	Sales Manager	NaN	NaN	NaN	NaN	Full-time	Santa Clarita, CA	NaN	NaN	FULL_TIME
33242	108965123	NaN	Office Administrative Assistant	NaN	NaN	NaN	NaN	Full-time	New York, NY	2.0	NaN	FULL_TIME
33243	102339515	52132271.0	Franchise Owner	NaN	NaN	NaN	NaN	Full-time	Greater Boston	NaN	NaN	FULL_TIME
33244	85008768	NaN	Licensed Insurance Agent	52000.0	NaN	45760.0	YEARLY	Full-time	Chico, CA	NaN	NaN	FULL_TIME
33245	3958427	630152.0	Stylist/ Clorist	80000.0	NaN	35000.0	YEARLY	Full-time	Chicago, IL	NaN	NaN	FULL_TIME

33246 rows × 12 columns

Missing values

In [7]:
```
# check the number of missing values for each column
new_data.isnull().sum()
```

Out[7]:
```
job_id                         0
company_id                   654
title                          0
max_salary                 22135
med_salary                 31005
min_salary                 22135
pay_period                 19894
formatted_work_type            0
location                       0
applies                    17008
formatted_experience_level  9181
work_type                      0
dtype: int64
```

First, simply drop missing values in the most important columns:

In [8]:
```
# drop missing values in columns 'company_id', 'max_salary', and 'min_salary'
new_data = new_data.dropna(subset=['company_id', 'max_salary', 'min_salary'])
new_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10919 entries, 4 to 33245
Data columns (total 12 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   job_id                      10919 non-null  int64
 1   company_id                  10919 non-null  float64
 2   title                       10919 non-null  object
 3   max_salary                  10919 non-null  float64
 4   med_salary                  0 non-null      float64
 5   min_salary                  10919 non-null  float64
 6   pay_period                  10919 non-null  object
 7   formatted_work_type         10919 non-null  object
 8   location                    10919 non-null  object
 9   applies                     6490 non-null   float64
 10  formatted_experience_level  8226 non-null   object
 11  work_type                   10919 non-null  object
dtypes: float64(5), int64(1), object(6)
memory usage: 1.1+ MB
```

Now we are left with 10,919 entries where these columns are non-null, but we still have null values in med_salary (median salary) and in some other columns as well.

- For `med_salary`, we can impute median salaries by taking the average of these maximum salary and minimum salary.
- `applies` represents the number of applications that have been submitted, it might be reasonable to impute 0 where this is missing, assuming a missing value indicates no applications were received or recorded.
- For `formatted_experience_level`, we can create a new category named "Not Specified" to prevent the loss of these records since they are substantial in number.

In [9]:
```
new_data = new_data.copy()

# Impute med_salary
new_data.loc[:, 'med_salary'] = (new_data['max_salary'] + new_data['min_salary']) / 2

# Fill `applies` with zeros where it is null
new_data.loc[:, 'applies'] = new_data['applies'].fillna(0)

# Creating a new category "Not Specified" for formatted_experience_level
new_data.loc[:, 'formatted_experience_level'] = new_data['formatted_experience_level'].fillna('Not Specified')

# After these operations, you can use new_data.info() to check the changes
new_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10919 entries, 4 to 33245
Data columns (total 12 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   job_id                      10919 non-null  int64
 1   company_id                  10919 non-null  float64
 2   title                       10919 non-null  object
 3   max_salary                  10919 non-null  float64
 4   med_salary                  10919 non-null  float64
 5   min_salary                  10919 non-null  float64
 6   pay_period                  10919 non-null  object
 7   formatted_work_type         10919 non-null  object
 8   location                    10919 non-null  object
 9   applies                     10919 non-null  float64
 10  formatted_experience_level  10919 non-null  object
 11  work_type                   10919 non-null  object
dtypes: float64(5), int64(1), object(6)
memory usage: 1.1+ MB
```

Normalized Salary

Add a column to show normalized salaries

In [10]:
```
# check pay_pariod column
new_data['pay_period'].value_counts()
```

Project Topic: Job market in the United States in 2023

Out[10]:
```
YEARLY     7265
HOURLY     3427
MONTHLY     144
WEEKLY       82
ONCE          1
Name: pay_period, dtype: int64
```

In [11]:
```
# Remove a entry with 'ONCE' in 'Pay Period'
new_data = new_data[new_data['pay_period'] != 'ONCE']

time_multiplier = {'HOURLY': 40*52, 'WEEKLY': 52, 'MONTHLY': 12, 'YEARLY': 1}

new_data['normalized_salary'] = new_data.apply(lambda x: (x['max_salary']+x['min_salary'])/2 if not pd.isnull(x['max_salary']) else x['med_salary'], axis=1)
new_data['normalized_salary'] = new_data.apply(
    lambda x: time_multiplier[x['pay_period']] * x['normalized_salary']
    if not pd.isna(x['pay_period']) and not pd.isna(x['normalized_salary'])
    else np.nan,
    axis=1
)
new_data = new_data.loc[new_data['pay_period'] != 'ONCE'].copy()
```

Split the location column into city and state

In [17]:
```
# Define a function to split the location into city and state
def split_city_state(location):
    parts = location.split(', ')
    if len(parts) == 2:  # If there are two parts, it's likely in the format "City, State"
        return parts[0], parts[1]
    else:
        return parts[0], None  # If not, return the whole location as city and None for state

# Apply the function to each location entry
new_data[['city', 'state']] = new_data.apply(lambda row: pd.Series(split_city_state(row['location'])), axis=1)

# Check the results
new_data
```

Out[17]:
```
array(['United States', "Coeur d'Alene, ID", 'Waukesha, WI', ...,
       'San Gabriel, CA', 'Arlington, WA', 'Winnetka, IL'], dtype=object)
```

Save to excel file

In [13]:
```
new_data.to_excel('job_postings_cleaned_2.xlsx', index=False)
```

Project Topic: Job market in the United States in 2023

Appendix 3: Categorization Process

Categorization

This notebook is trying to categorize the industry into 3 sectors

In [1]:
```
import pandas as pd
import numpy as np

job_posting = pd.read_excel("job_postings_cleaned_1.xlsx")
job_industry = pd.read_csv("job_details/job_industries.csv")
industry = pd.read_csv("maps/industries.csv")
```

In [8]:
```
# Merge to create industry_name_map
industry_name_map = job_posting.merge(job_industry, on="job_id", how="left").merge(industry, on="industry_id", how="left")

# New industry names data, referencing industry code v2 standard
new_industry_data = {
    'industry_id': [21, 66, 72, 107, 108, 923, 1221, 1359, 1594, 1905, 1965, 2226, 3102, 3208, 3216, 3219, 3221],
    'industry_name': [
        'Tobacco Manufacturing', 'Fisheries', 'Legislative Offices', 'Political Organizations',
        'Translation and Localization', 'HVAC and Refrigeration Equipment Manufacturing',
        'Wholesale Drugs and Sundries', 'Retail Health and Personal Care Products',
        'Technology, Information and Media', 'Holding Companies', 'Janitorial Services',
        'Vehicle Repair and Maintenance', 'IT System Custom Software Development', 'E-Learning',
        'Industrial Automation', 'Commercial Real Estate', 'Mechanical or Industrial Engineering'
    ]
}

# Create a DataFrame from the new industry data
new_industries_df = pd.DataFrame(new_industry_data)

# Merge the new industry names into the existing map
industry_name_map = industry_name_map.merge(new_industries_df, on="industry_id", how="left")

# Update the industry_name with the new names where available
industry_name_map['industry_name'] = industry_name_map['industry_name_y'].fillna(industry_name_map['industry_name_x'])

# Drop the temporary columns
industry_name_map.drop(['industry_name_x', 'industry_name_y'], axis=1, inplace=True)
```

In [9]:
```
industry_name_map
```

	title	max_salary	med_salary	min_salary	pay_period	formatted_work_type	location	applies	formatted_experience_level	work_type	normalized_salary	city	state	industry_id	industry_name
	Sales Manager	350000.0	237500.0	125000.0	YEARLY	Full-time	Coeur d'Alene, ID	0	Mid-Senior level	FULL_TIME	237500.0	Coeur d'Alene	ID	43.0	Financial Services
	Body Technician	30.0	25.0	20.0	HOURLY	Full-time	Waukesha, WI	0	Entry level	FULL_TIME	52000.0	Waukesha	WI	3198.0	Automotive
	Controls Engineer	150000.0	135500.0	121000.0	YEARLY	Full-time	Orlando, FL	1	Mid-Senior level	FULL_TIME	135500.0	Orlando	FL	135.0	Industrial Machinery Manufacturing
	CDL Class B Driver	27.0	26.0	25.0	HOURLY	Full-time	Oakland, CA	0	Mid-Senior level	FULL_TIME	54080.0	Oakland	CA	92.0	Truck Transportation
	Senior Manager, Indirect Procurement	143000.0	119500.0	96000.0	YEARLY	Full-time	Broomfield, CO	0	Mid-Senior level	FULL_TIME	119500.0	Broomfield	CO	23.0	Food and Beverage Manufacturing
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Human Resources Senior Manager	180000.0	150000.0	120000.0	YEARLY	Full-time	Savannah, GA	1	Not Specified	FULL_TIME	150000.0	Savannah	GA	104.0	Staffing and Recruiting
	Registered Nurse (RN) Vaccinator	50.0	50.0	50.0	HOURLY	Part-time	Muskegon, MI	0	Not Specified	PART_TIME	104000.0	Muskegon	MI	104.0	Staffing and Recruiting
	Office Associate	42000.0	39500.0	37000.0	YEARLY	Full-time	Albany, GA	5	Not Specified	FULL_TIME	39500.0	Albany	GA	27.0	Retail
	Office Associate	42000.0	39500.0	37000.0	YEARLY	Full-time	Albany, GA	5	Not Specified	FULL_TIME	39500.0	Albany	GA	34.0	Food and Beverage Services
	Stylist/ Clorist	80000.0	57500.0	35000.0	YEARLY	Full-time	Chicago, IL	0	Not Specified	FULL_TIME	57500.0	Chicago	IL	18.0	Personal Care Product Manufacturing

In [10]:
```
industry_name_map["industry_name"].isnull().sum()
```

Out[10]: 8

In [16]:
```
# Defining the mapping of industry names to categories based on the given categories
category_mapping = {
    'Technology, Engineering, and Science': [
        'Industrial Machinery Manufacturing', 'Software Development', 'Biotechnology Research',
        'IT Services and IT Consulting', 'Telecommunications', 'Semiconductor Manufacturing',
        'Computer Hardware Manufacturing', 'Computers and Electronics Manufacturing',
        'Civil Engineering', 'Chemical Manufacturing', 'Environmental Services',
        'Biotechnology', 'Computer Games', 'Data Science Program', 'Electrical and Computer Engineering Department',
        'Robotics Engineering Department', 'Bioinformatics and Computational Biology Program',
        'Physics Department', 'Computer Science Department', 'Aerospace Engineering Department',
        'Technology, Information and Internet', 'Automotive', 'Aviation & Aerospace',
        'Aviation and Aerospace Component Manufacturing', 'Defense & Space', 'Defense and Space Manufacturing',
        'Electrical & Electronic Manufacturing', 'Energy', 'Engineering Services', 'Manufacturing',
        'Mechanical Or Industrial Engineering', 'Mining', 'Mining & Metals', 'Oil & Energy', 'Oil and Gas',
        'Renewable Energy Equipment Manufacturing', 'Renewable Energy Semiconductor Manufacturing',
        'Renewables & Environment', 'Solar Electric Power Generation', 'Utilities',
        'Wireless Services', 'Electric Power Generation', 'Renewable Energy Power Generation',
        'Semiconductor Manufacturing', 'Alternative Energy Sources', 'Nanotechnology Research',
        'Information Technology & Services', 'Shipbuilding', 'Automation Machinery Manufacturing',
        'Appliances, Electrical, and Electronics Manufacturing', 'Sporting Goods Manufacturing',
        'Furniture and Home Furnishings Manufacturing', 'Medical Equipment Manufacturing',
        'Pharmaceutical Manufacturing', 'Research Services', 'Services for Renewable Energy',
        'Motor Vehicle Manufacturing', 'Airlines and Aviation', 'Textile Manufacturing',
        'Computer Networking Products', 'Plastics Manufacturing', 'Computer and Network Security',
        'Machinery Manufacturing', 'Packaging and Containers Manufacturing', 'Electric Power Transmission, Control, and Distribution',
        'Maritime Transportation', 'Equipment Rental Services', 'Warehousing and Storage',
        'Glass, Ceramics and Concrete Manufacturing', 'Internet Marketplace Platforms',
        'Movies, Videos, and Sound', 'Printing Services', 'IT System Custom Software Development',
        'Railroad Equipment Manufacturing', 'Tobacco Manufacturing', 'Technology, Information and Media'
        # Additional industries in this category should be added here
    ],
    'Business, Finance, and Legal Services': [
        'Financial Services', 'Banking', 'Insurance', 'Accounting', 'Law Practice',
        'Legal Services', 'Management Consulting', 'Investment Banking', 'Real Estate',
        'Advertising Services', 'Marketing Services', 'Business School', 'Business Consulting and Services',
        'Human Resources Services', 'Investment Management', 'Economic Programs', 'Commercial Real Estate',
```

Project Topic: Job market in the United States in 2023


```
        'Consulting', 'International Trade & Development', 'International Trade and Development',
        'Investment Banking/Venture', 'Market Research', 'Professional Training & Coaching',
        'Public Policy', 'Staffing & Recruiting', 'Venture Capital & Private Equity',
        'Wholesale', 'Capital Markets', 'Professional Services', 'Public Relations & Communications',
        'Public Relations and Communications Services', 'Retail', 'Transportation/Trucking/Railroad',
        'Management Consulting Services', 'Executive Offices', 'Leasing Non-residential Real Estate',
        'Venture Capital and Private Equity Principals', 'Strategic Management Services',
        'Truck Transportation', 'Staffing and Recruiting', 'Outsourcing and Offshoring Consulting',
        'Transportation, Logistics, Supply Chain and Storage', 'Facilities Services',
        'Wholesale Food and Beverage', 'Architecture and Planning', 'Information Services',
        'Paper and Forest Product Manufacturing', 'Wholesale Building Materials',
        'Security and Investigations', 'Newspaper Publishing', 'Residential Building Construction',
        'Furniture', 'Administration of Justice', 'Philanthropic Fundraising Services',
        'Administrative and Support Services', 'Telephone Call Centers', 'Professional Training and Coaching',
        'International Affairs', 'Human Resources', 'Legislative Offices', 'Veterinary',
        'Non-profit Organization Management', 'Business Supplies & Equipment', 'Janitorial Services',
        'Building Materials'
        # Additional industries in this category should be added here
    ],
    'Arts, Humanities, Social Sciences, and Interdisciplinary Studies': [
        'Media Production', 'Entertainment', 'Design Services', 'Education Administration Programs',
        'Government Administration', 'Public Relations and Communications Services',
        'Humanities and Arts Department', 'Social Science and Policy Studies Department',
        'Interdisciplinary Programs', 'Animation and Post-production', 'Writing and Editing',
        'Journalism', 'Publishing', 'Book and Periodical Publishing', 'Non-profit Organizations',
        'Civic and Social Organizations', 'Public Safety', 'Libraries', 'Higher Education',
        'Museums & Institutions', 'Museums, Historical Sites, and Zoos', 'Performing Arts',
        'Spectator Sports', 'Broadcast Media', 'Broadcast Media Production and Distribution',
        'Entertainment Providers', 'Hospitality', 'Travel Arrangements', 'Food and Beverage Services',
        'Food & Beverages','Education Management', 'Government Relations Services', 'Artists and Writers',
        'Public Policy Offices', 'Primary and Secondary Education', 'Primary/Secondary Education',
        'Restaurants', 'Retail Apparel and Fashion', 'Retail Art Supplies', 'Retail Groceries',
        'Retail Luxury Goods and Jewelry', 'Retail Motor Vehicles', 'Retail Office Equipment',
        'Consumer Services', 'Events Services', 'Food Production', 'Individual and Family Services',
        'Online Media', 'Recreational Facilities', 'Recreational Facilities & Services',
        'Religious Institutions', 'Think Tanks', 'Household Services', 'Personal Care Product Manufacturing',
        'Consumer Electronics', 'Consumer Goods', 'Cosmetics', 'E-Learning Providers', 'Educational Services',
        'Leisure, Travel & Tourism', 'Online Audio and Video Media', 'Hospital & Health Care',
        'Medical Practices', 'Apparel & Fashion', 'Luxury Goods & Jewelry', 'Musicians',
        'Photography', 'Graphic Design', 'Health, Wellness & Fitness', 'Wellness and Fitness Services',
        'Food and Beverage Manufacturing', 'Dairy Product Manufacturing', 'Hospitals and Health Care',
        'Mental Health Care', 'Construction', 'Gambling Facilities and Casinos',
        'Physical, Occupational and Speech Therapists', 'Veterinary Services',
        'Alternative Dispute Resolution', 'Farming', 'Ranching', 'Beverage Manufacturing',
        'Alternative Medicine', 'Research', 'Fundraising', 'Freight and Package Transportation',
        'Law Enforcement', 'Education', 'Home Health Care Services', 'Armed Forces',
        'Ground Passenger Transportation', 'Wholesale Import and Export', 'Dentists',
        'Political Organizations', 'Translation and Localization', 'Education', 'Veterinary',
        'Vehicle Repair and Maintenance', 'Urban Transit Services', 'Rail Transportation'
        # Additional industries in this category should be added here
    ],
    'Other': [
        # Industries that do not fit into the above categories
    ]
}


# Function to categorize industry
def categorize_industry(industry_name):
    for category, industries in category_mapping.items():
        if industry_name in industries:
            return category
    return 'Other'  # Default category if not found

# Apply the function to create a new column
industry_name_map['categorize'] = industry_name_map['industry_name'].apply(categorize_industry)

industry_name_map  # Displaying the DataFrame for confirmation
```

	max_salary	med_salary	min_salary	pay_period	formatted_work_type	location	applies	formatted_experience_level	work_type	normalized_salary	city	state	industry_id	industry_name	categorize
5 r	350000.0	237500.0	125000.0	YEARLY	Full-time	Coeur d'Alene, ID	0	Mid-Senior level	FULL_TIME	237500.0	Coeur d'Alene	ID	43.0	Financial Services	Business, Finance, and Legal Services
y n	30.0	25.0	20.0	HOURLY	Full-time	Waukesha, WI	0	Entry level	FULL_TIME	52000.0	Waukesha	WI	3198.0	Automotive	Technology, Engineering, and Science
5 r	150000.0	135500.0	121000.0	YEARLY	Full-time	Orlando, FL	1	Mid-Senior level	FULL_TIME	135500.0	Orlando	FL	135.0	Industrial Machinery Manufacturing	Technology, Engineering, and Science
3 r	27.0	26.0	25.0	HOURLY	Full-time	Oakland, CA	0	Mid-Senior level	FULL_TIME	54080.0	Oakland	CA	92.0	Truck Transportation	Business, Finance, and Legal Services
r , t t	143000.0	119500.0	96000.0	YEARLY	Full-time	Broomfield, CO	0	Mid-Senior level	FULL_TIME	119500.0	Broomfield	CO	23.0	Food and Beverage Manufacturing	Arts, Humanities, Social Sciences, and Interdi...
.	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
n s r r	180000.0	150000.0	120000.0	YEARLY	Full-time	Savannah, GA	1	Not Specified	FULL_TIME	150000.0	Savannah	GA	104.0	Staffing and Recruiting	Business, Finance, and Legal Services
d) r	50.0	50.0	50.0	HOURLY	Part-time	Muskegon, MI	0	Not Specified	PART_TIME	104000.0	Muskegon	MI	104.0	Staffing and Recruiting	Business, Finance, and Legal Services
9 9	42000.0	39500.0	37000.0	YEARLY	Full-time	Albany, GA	5	Not Specified	FULL_TIME	39500.0	Albany	GA	27.0	Retail	Business, Finance, and Legal Services
9 9	42000.0	39500.0	37000.0	YEARLY	Full-time	Albany, GA	5	Not Specified	FULL_TIME	39500.0	Albany	GA	34.0	Food and Beverage Services	Arts, Humanities, Social Sciences, and Interdi...
/ t	80000.0	57500.0	35000.0	YEARLY	Full-time	Chicago, IL	0	Not Specified	FULL_TIME	57500.0	Chicago	IL	18.0	Personal Care Product Manufacturing	Arts, Humanities, Social Sciences, and Interdi...

In [17]:
```
industry_name_map.to_excel('job_posting_industry.xlsx', index=False)
```

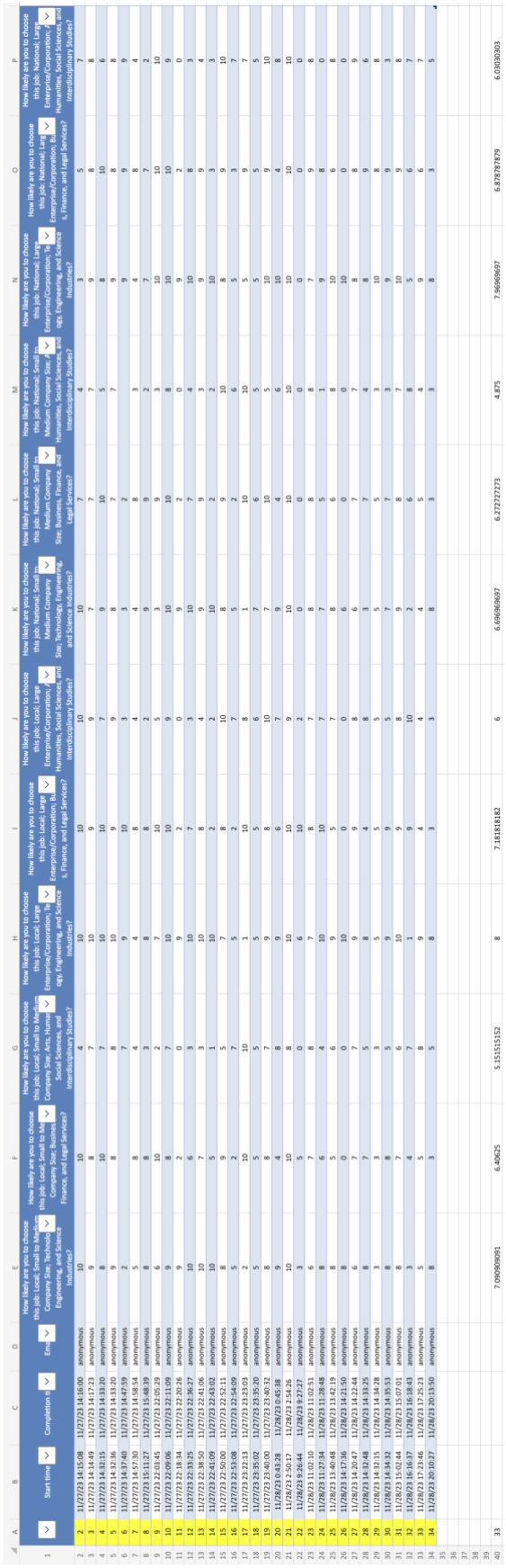
Appendix 4: Conjoint Analysis

Rating-Based Conjoint Analysis Survey Link

https://forms.office.com/Pages/ResponsePage.aspx?id=9XacWBXK-UGIS1XsFaBnKgUHn9DCM9NNjyNk1uq9qfRUQVRLRE9BRUZCTlQwQUtMVkRZMTJTRDJMRS4u

Project Topic: Job market in the United States in 2023

Regression Analysis Process



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		Start time	Completion time	Email	How likely are you to choose this job: Local; Small to Medium Company Size; Technology, Engineering, and Science Holdings?	How likely are you to choose this job: Local; Small to Medium Company Size; Business, Finance, and Legal Services?	How likely are you to choose this job: Local; Small to Medium Company Size; Arts, Humanities, Social Sciences, and Interdisciplinary Studies?	How likely are you to choose this job: Local; Large Enterprise/Corporation; Technology, Engineering, and Science Holdings?	How likely are you to choose this job: Local; Large Enterprise/Corporation; Finance, and Legal Services?	How likely are you to choose this job: Local; Large Enterprise/Corporation; Humanities, Social Sciences, and Interdisciplinary Studies?	How likely are you to choose this job: National; Small to Medium Company Size; Technology, Engineering, and Science Holdings?	How likely are you to choose this job: National; Small to Medium Company Size; Business, Finance, and Legal Services?	How likely are you to choose this job: National; Small to Medium Company Size; Humanities, Social Sciences, and Interdisciplinary Studies?	How likely are you to choose this job: National; Large Enterprise/Corporation; Technology, Engineering, and Science Holdings?	How likely are you to choose this job: National; Large Enterprise/Corporation; Finance, and Legal Services?	How likely are you to choose this job: National; Large Enterprise/Corporation; Humanities, Social Sciences, and Interdisciplinary Studies?
2	2	11/27/23 14:15:08	11/27/23 14:16:00	anonymous	10	10	4	10	10	10	10	7	7	3	5	7
3	3	11/27/23 14:14:49	11/27/23 14:17:23	anonymous	9	8	7	9	9	9	7	7	7	9	8	8
4	4	11/27/23 14:32:15	11/27/23 14:33:20	anonymous	8	10	7	10	10	7	9	10	5	8	10	6
5	5	11/27/23 14:32:36	11/27/23 14:33:20	anonymous	9	8	8	10	9	9	8	7	7	9	8	8
6	6	11/27/23 14:37:40	11/27/23 14:47:59	anonymous	2		7	9	10	3	3	2		9	9	9
7	7	11/27/23 14:57:30	11/27/23 14:58:54	anonymous	5	8	4	4	8	4	4	8	3	4	8	4
8	8	11/27/23 15:11:27	11/27/23 15:48:39	anonymous	8	8	3	8	8	2	9	9	2	7	7	2
9	9	11/27/23 22:03:45	11/27/23 22:05:29	anonymous	6	10	2	7	10	5	3	9	3	10	10	10
10	10	11/27/23 22:09:48	11/27/23 22:11:09	anonymous	9	8	7	10	10	9	10	10	8	10	10	9
11	11	11/27/23 22:18:34	11/27/23 22:20:26	anonymous	9	2	0	9	2	0	9	2	0	9	2	0
12	12	11/27/23 22:33:25	11/27/23 22:36:27	anonymous	10	6	3	10	7	3	10	7	4	10	8	3
13	13	11/27/23 22:38:50	11/27/23 22:41:06	anonymous	10	7	3	10	8	4	9	9	3	9	9	4
14	14	11/27/23 22:41:09	11/27/23 22:43:02	anonymous	10	5	1	10	2	2	10	2	2	10	3	3
15	15	11/27/23 22:50:00	11/27/23 22:52:11	anonymous	8	9	5	7	8	10	8	9	10	8	9	10
16	16	11/27/23 22:53:08	11/27/23 22:54:09	anonymous	5	2	7	5	2	7	5	2	6	5	3	7
17	17	11/27/23 23:33:23	11/27/23 23:35:20	anonymous	2	10	10	1	10	8	1	10	10	5	9	7
18	18	11/27/23 23:35:02	11/27/23 23:40:32	anonymous	5	5	5	5	8	6	7	6	5	5	5	5
19	19	11/27/23 23:40:00	11/28/23 0:43:28	anonymous	8	8	7	9	8	10	7	10	5	10	9	10
20	20	11/28/23 0:43:28	11/28/23 2:50:17	anonymous	9	4	8	9	6	7	9	4	6	10	4	8
21	21	11/28/23 2:50:17	11/28/23 2:54:26	anonymous	10	10	8	10	10	9	10	10	10	10	10	10
22	22	11/28/23 9:26:44	11/28/23 9:27:27	anonymous	3	5	0	6	10	2	0	0	0	0	0	0
23	23	11/28/23 11:02:10	11/28/23 11:02:51	anonymous	6	7	8	7	8	7	8	8	8	7	9	8
24	24	11/28/23 11:27:34	11/28/23 11:28:48	anonymous	8	6	4	10	10	7	7	5	1	9	8	0
25	25	11/28/23 13:47:19	11/28/23 13:48:15	anonymous	8	5	6	9	5	7	8	6	8	10	6	8
26	26	11/28/23 14:17:36	11/28/23 14:21:50	anonymous	8	0	6	10	0	0	6	0	0	10	6	0
27	27	11/28/23 14:20:47	11/28/23 14:22:44	anonymous	6	7	7	9	9	8	6	7	7	8	8	9
28	28	11/28/23 14:32:48	11/28/23 14:33:25	anonymous	8	7	5	8	4	8	3	7	4	8	9	6
29	29	11/28/23 14:32:15	11/28/23 14:34:28	anonymous	3	3	3	5	5	5	5	5	3	10	8	8
30	30	11/28/23 14:34:32	11/28/23 14:35:53	anonymous	8	8	5	9	9	5	7	7	3	9	9	3
31	31	11/28/23 15:02:44	11/28/23 15:07:01	anonymous	8	7	6	10	9	8	9	8	7	10	9	8
32	32	11/28/23 16:16:37	11/28/23 16:18:43	anonymous	3	4	7	1	9	10	2	6	8	5	6	7
33	33	11/28/23 21:23:46	11/28/23 21:25:23	anonymous	5	5	8	9	4	4	4	5	4	9	6	7
34	34	11/28/23 20:10:27	11/28/23 20:13:50	anonymous	8	3	5	8	3	3	8	3	3	8	3	5
35																
36																
37																
38																
39																
40	33				7.090909091	6.40625	5.151515152	8	7.181818182	6	6.696969697	6.272727273	4.875	7.969696697	6.878787879	6.03030303

Project Topic: Job market in the United States in 2023

	A	B	C	D	E	F
1		AVERAGE	Local	Small to Medium Company Size	Technology, Engineering, and Science Industries	Business, Finance, and Legal Services
2	Local; Small to Medium Company Size; Technology, Engineering, and Science Industries	7.091	1	1	1	0
3	Local; Small to Medium Company Size; Business, Finance, and Legal Services	6.406	1	1	0	1
4	Local; Small to Medium Company Size; Arts, Humanities, Social Sciences, and Interdisciplinary Studies	5.152	1	1	0	0
5	Local; Large Enterprise/Corporation; Technology, Engineering, and Science Industries	8.000	1	0	1	0
6	Local; Large Enterprise/Corporation; Business, Finance, and Legal Services	7.182	1	0	0	1
7	Local; Large Enterprise/Corporation; Arts, Humanities, Social Sciences, and Interdisciplinary Studies	6.000	1	0	0	0
8	National; Small to Medium Company Size; Technology, Engineering, and Science Industries	6.697	0	1	1	0
9	National; Small to Medium Company Size; Business, Finance, and Legal Services	6.273	0	1	0	1
10	National; Small to Medium Company Size; Arts, Humanities, Social Sciences, and Interdisciplinary Studies	4.875	0	1	0	0
11	National; Large Enterprise/Corporation; Technology, Engineering, and Science Industries	7.970	0	0	1	0
12	National; Large Enterprise/Corporation; Business, Finance, and Legal Services	6.879	0	0	0	1
13	National; Large Enterprise/Corporation; Arts, Humanities, Social Sciences, and Interdisciplinary Studies (Benchmark)	6.030	0	0	0	0

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.992367592							
5	R Square	0.984793438							
6	Adjusted R Square	0.976103974							
7	Standard Error	0.150097613							
8	Observations	12							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	4	10.21315038	2.553287596	113.3318983	1.92821E-06			
13	Residual	7	0.157705054	0.022529293					
14	Total	11	10.37085544						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17	Intercept	5.885890152	0.096887593	60.74967899	8.59819E-11	5.6567874	6.114992903	5.6567874	6.114992903
18	Local	0.184501263	0.086658897	2.129051584	0.070769241	-0.020414468	0.389416993	-0.020414468	0.389416993
19	Small to Medium Company Size	-0.927872475	0.086658897	-10.70718072	1.3612E-05	-1.132788205	-0.722956745	-1.132788205	-0.722956745
20	Technology, Engineering, and Science Industries	1.925189394	0.10613504	18.13905561	3.82845E-07	1.674219904	2.176158884	1.674219904	2.176158884
21	Business, Finance, and Legal Services	1.170691288	0.10613504	11.03020536	1.11703E-05	0.919721798	1.421660778	0.919721798	1.421660778
22									
23									
24	US Location	0.184501263	6%						
25	Company Size	0.927872475	31%						
26	Industry Category	1.925189394	63%						
27									
28		3.037563131	100%						