Table of Contents

Overview

The *My Career Advisor (MCA)* is a digital career guidance tool designed to support Filipino Senior High School learners in identifying job opportunities that align with their skills, interests, values, and future education plans. The platform integrates multiple assessment sections—including College Plans, Passions, Aptitudes, Interests, and Work Values—to build a personalized learner profile and recommend career paths accordingly.

MCA generates two types of job recommendation lists: (1) one based solely on learners' self-identified passions, and (2) a second, more comprehensive list derived from their aptitude, interest, and value assessments. College plans serve as a filtering input across both lists to ensure alignment with educational requirements for job roles.

The design of each assessment section is grounded in research from the fields of psychology, career education, and future orientation theory. The tool is built with flexibility in mind: learners or institutions can choose between Questionnaire (Likert-style) and Slider-based formats depending on their context. To ensure transparency, MCA provides reporting features for each section, and validation of outputs is conducted at both the industry and DepEd levels.

System Overview

Description:
My Career Advisor (MCA) is a digital self-assessment tool designed to help Filipino senior high school students explore their career options. The platform analyzes a learner's inputs related to **College Plans**, **Passions**, **Aptitudes**, **Interests**, and **Work Values**, and generates a **Career Pathway Report** listing matched job roles.

Primary Use Case:
To guide learners in identifying suitable career paths aligned with their personal and contextual attributes.

Intended Use

Intended Use
→ Main use for MCA Philippines

→ Other possible, interoperable, uses with job list and educational tagging

Background

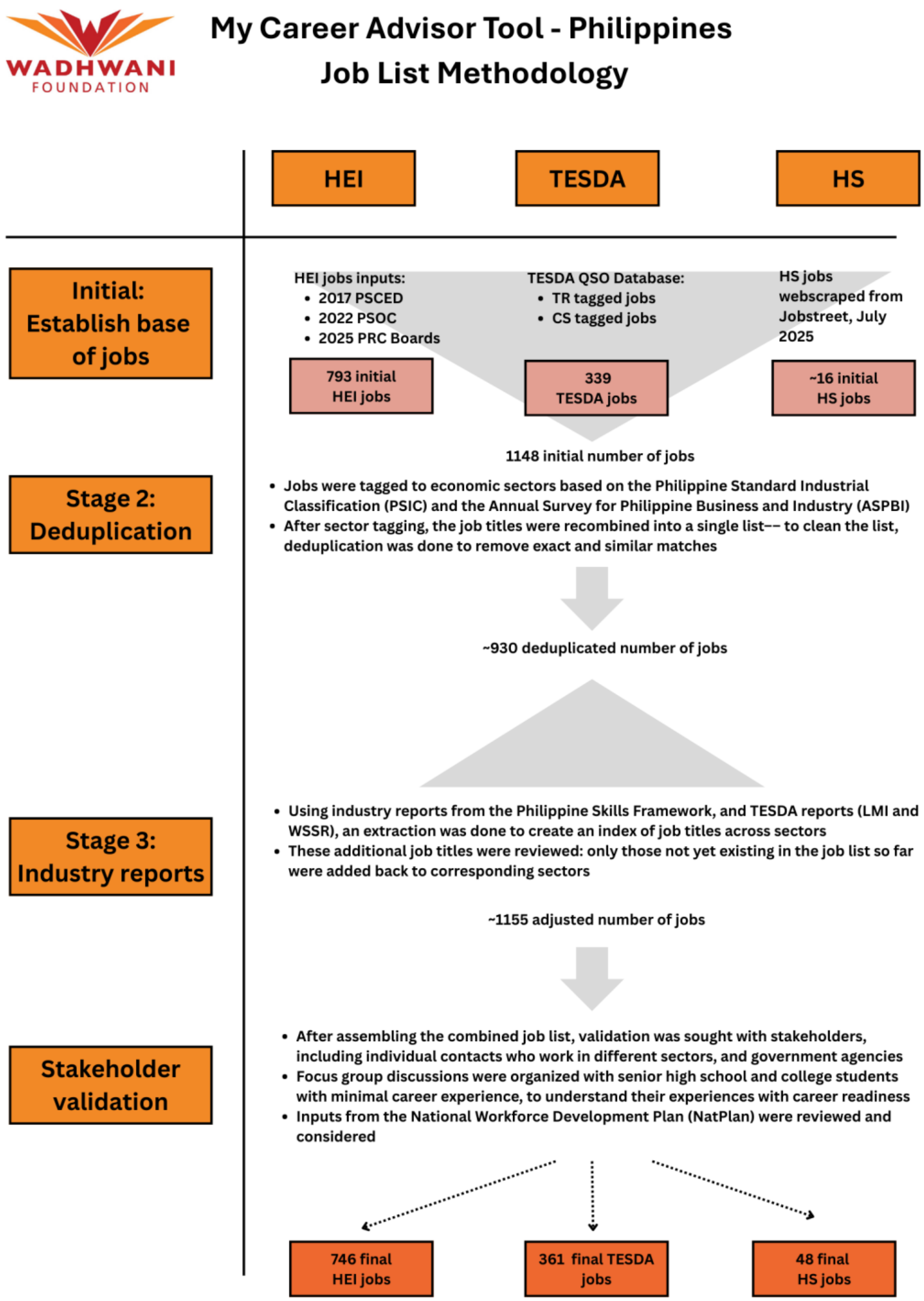
My Career Advisor (MCA) Tool in India

The Wadhwani Foundation launched the MCA tool with Ministry of Education India on July 30th, 2025 for national roll-out. Learn more about the launch here: https://www.youtube.com/watch?v=Tcz6YIxkCuk

Rational for Developing a PH Job List

The MCA Job List was not adapted from a generic global dataset—it was built deliberately for the Philippines. To scale with government partners like DepEd, CHED, and TESDA, we committed to creating a job list that reflects actual education pathways available to Filipino learners and real labor market demand in the local economy. That meant going beyond international taxonomies to integrate the country's own PSCED (CHED programs), TESDA NC qualifications, and DOLE sector classifications. By linking these education credentials to current job roles and validating them with employers, government stakeholders, and youth, the MCA Job List becomes more than a career guidance tool—it becomes a bridge between schooling and work. This local specificity ensures the tool is not just usable, but impactful—positioned for public-private deployment at scale and responsive to the evolving priorities of the Philippine skilling ecosystem.

Executive Summary



Output Table Schema for My Career Advisor Tool

The following table shows the final output created from the inputs, for the purpose of a useable database for the localized Philippine version of the My Career Advisor tool. Each entry is made up of a job title, the sector and subsector it is tagged to, and the minimum education required if higher education, or no higher education needed.

Job Title	Sector	Subsector	Education Required

Breakdown of job titles across education pathways

Education Level	Number of Jobs
Higher Education Institution – Professional Regulatory Commission (HEI PRC)	583
Higher Education Institution – Non-Licensed (HEI Non-PRC)	163
TESDA National Certificate (TESDA NC)	361
Senior High School (SHS) Diploma	48
TOTAL	1155

To assemble at least 1000 different jobs that capture the Philippine labor market, it was essential to utilize different sources. Job titles were segmented into 4 main education segments to break down the total list, and identify different ways and inputs to obtain the jobs for each:

- HEI PRC (Jobs tagged to higher education degrees with PRC licenses)
- HEI Non-PRC (Jobs tagged to higher education degrees without PRC licenses)
- TESDA NC (Jobs tagged to TESDA national certificate programs)
- HS (Jobs tagged to high school diplomas)

for HEI-PRC

For HEI PRC jobs, the Republic of the Philippines Professional Regulation Commission (PRC) hosts the updated list of professions with licenses. These include professions that require graduating from specific higher education degrees, among other requirements. The graduates must take a board exam governed by one of the 45 active boards as of 2025. By consulting this method and reviewing the specific qualifications from the PRC website, titles tied to higher education degrees and PRC licenses can be extracted. Some job titles are tied to post-bachelor's or master's level education, excluding them from the final list of entry level jobs.

for HEI-non PRC

For HEI Non-PRC jobs, the 2017 Philippine Standard Classification of Education (PSCED) and 2022 Updates to the 2012 Philippine Standard Occupational Classification (PSOC) were utilized to generate job titles tagged to bachelor level degree programs. There is no public database that links degree programs to specific kinds of jobs. Because of this, the PSOC was consulted to identify likely jobs that would correspond to bachelor level degree programs. Prompting a large-language AI model, an index of the updated 2022 PSOC job titles was uploaded alongside an index of 2017 PSCED bachelor level degree programs. AI was tasked to inspect the PSOC titles and to assign a likely job title for each unique qualification. From this assignment, several rounds of inspection, title reiteration, and deduplication of similar titles were done to arrive at an updated list of HEI Non-PRC tagged jobs.

TESDA National Certificate (NC) jobs

For TESDA NC jobs, the foundation team had access to an internal database by the TESDA Competency Standards Development Division (CSSD-QSO) that linked TESDA NCs to specific job titles. The list of job titles tagged to NCs were extracted, cleaned, and deduplicated for multiples entries that represented single job functions albeit with different skill components.

high school jobs

For HS jobs, a Philippine Business for Education study was first consulted before conducting a webscrape of Philippine jobsites to obtain a sample of jobs tagged to high school diplomas. From a volume of jobs with different keywords, frequently repeating words and synonyms were identified to cluster and reduce the entries. The final number of high school jobs is not many compared to technical vocational educational training and higher education jobs.

how to get job sectors

Across all these education segments, the 2019 Updates to the 2009 Philippine Standard Industrial Classification (PSIC) and 2022 Annual Survey of Philippine Business and Industry (ASPBI) were used to establish a method to tag jobs to sectors and subsectors. Jobs across educational segments were tagged to sectors. Subsectors were added at the end of the process, adding a level of granularity to the labor market segments that are useful for generating descriptions of jobs by the product team.

validation with PSF and EDCOM

To supplement the job titles obtained across the segments and ensure industry relevance, research was done to obtain and review sector-based reports to extract applicable job titles to add back to their corresponding sectors. Philippine Skills Frameworks for the Information, Communications, and Technology Industry (ICT) and Human Capital were available online. Meanwhile, two groups of publications from TESDA were consulted: TESDA Workplace Skills and Satisfaction Survey Reports (2020-2024) and TESDA Sector Labor Market Information reports. These reports provide rich insight into real job titles that reflect the Philippine labor market, created from consulting industry groups.

To triangulate the validation done for job titles generated from the educational and industry related inputs, government validation was sought after. The Second Congressional Commission on Education (EDCOM 2) leads efforts for the National Workforce Development Plan (NatPlan), a cross-agency cabinet cluster initiative that harmonizes policy and programs that bridge education to employment. Through dialogue with EDCOM 2's NatPlan team and other colleagues, the team was able to identify strategic areas for integrating the MCA and the foundation's broader work into the national government's executive programming for workforce development.

Lastly, keeping in mind the end-users, a youth validation step was done last to situate the usefulness of the tool in the lived experiences of Filipino young people. 3 focus group discussions (FGDs) were organized online with senior high school and college students, and 1 live poll in-person through a youth organizing network. These FGDs were not user tests of a demo or live tool, but rather probed learners' values, agency and context in accessing career readiness resources. A sample of the jobs were shown to the learners to obtain additional comments about the list and how they might be perceived by users once released.

Summary of Inputs Used

Input Sources	Shorthand	Outputs
2017 Philippine Standard Classification of Education (PSCED)	PSCED	Job Title, Education Required
2022 Updates to the 2012 Philippine Standard Occupational Classification (PSOC)	PSOC	Job Title
2019 Updates to the 2009 Philippine Standard Industrial Classification (PSIC)	PSIC	Sector, Subsector
2022 Annual Survey of Philippine Business and Industry (ASPBI)	ASPBI	Sector, Subsector
TESDA QSO Competency Bank Database	TESDA jobs	Job Title, Education Required
Webscraped job titles from job search websites	SHS jobs	Job Title
Republic of the Philippines Professional Regulation Commission (PRC) Boards	HEI PRC	Job Title
Philippine Skills Frameworks for the ICT Industry (DICT)	PSF	Job Title
Philippine Skills Framework – Human Capital	PSF	Job Title
TESDA Workplace Skills and Satisfaction Survey Reports (2020-2024)	TESDA WSSS	Job Title
TESDA Sector Labor Market Information 2022	TESDA LMI	Job Title
National Workforce Development Plan (lead by EDCOM 2, with CHED, TESDA, DepEd)	NatPlan	Job Title

Breakdown of Inputs

PSOC and PSCED is for HEI-non PRIC

Philippine Standard Classification of Education (PSCED)

Purpose: Classifies all educational programs in the Philippines by level and field of study
Use: Potential master list of bachelor level degree programs that serve as the basis for HEI job generation
Structure: 1 field, 3-4 hierarchical levels:

- **Level (1-digit)** – Broad education level (e.g., secondary, bachelor's, master's)
- **Broad Field (2-digits)** – General domain of study (e.g., Education, Engineering, Business)
- **Narrow Field (1-digit)**
- **Detailed Field (1 digit)** – Specific subfield or discipline (degree programs offered)
- Example**:**
    - **6** – Bachelor's or equivalent level *(note: level is different from field)*
        - **7** – Education field
            - **76** – Teacher Training with Subject Specialization
                - **762** – Teacher Training in Science and Math
                    - **76221 – Bachelor of Secondary Education – Major in Biology**

Philippine Standard Occupational Classification (PSOC)

Purpose: Classifies **jobs/occupations** based on the tasks individuals perform, from International Labour Organization (ILO) standards
Use: Government standard for delineating between jobs that require and do not require HEI degrees, according to skill levels.

Structure: 4 hierarchical levels:

- **Major Group** (1-digit)
- **Sub-major** (2-digit)
- **Minor** (3-digit)
- **Unit Group** (4-digit)

- Adapted from the International Standard Classification of Occupations (ISCO)
- Example:
    - 2 – Professionals
        - 25 – Information and Communications Technology (ICT) Professionals
            - 251 – Software and Applications Developers and Analysts
                - 2512 – Software Developers

Philippine Standard Industrial Classification (PSIC)

for the subsectors
by law, to start a business, you need to have a PSIC

Purpose: Classifies **businesses and establishments** based on their economic activity.
Use: Government standard for labor market driven sector and subsector tagging

Structure: 5 hierarchical levels:

    - **Sections** (1-letter code, Sectors)
    - **Divisions** (2-digit, Sub-Sectors)
    - **Groups** (3-digit)
    - **Classes** (4-digit)
    - **Sub-classes** (5-digit)
- Example:
- Section J – Information and Communication
    - J62 – Computer programming, consultancy and related activities
        - J620 – Computer programming, consultancy and related activities
            - J6201 – Computer programming activities
                - J62011 – Writing, modifying, testing software to meet user needs

List of Economic Sectors from PSIC

Sectors	Number of Subsectors
Accommodation and food service activities	5
Administrative and support service activities	19
Agriculture, forestry and fishing	10
Arts, entertainment and recreation	6
Construction	8
Education	6
Electricity, gas, steam and air conditioning supply	1
Financial and insurance activities	9
Human health and social work activities	9
Information and communication	13
Manufacturing	67
Mining and quarrying	6
Other service activities	6
Professional, scientific and technical activities	15
Public Administration and Defense; Compulsory Social Security	X
Real estate activities	2
Transportation and storage	9
Water supply; sewerage, waste management and remediation activities	6
Wholesale & retail trade; repair of motor vehicles & motorcycles	19

Note: Subsector count comes from the Annual Survey on Philippine Business and Industry (ASPBI). "Public administration and defense; compulsory social security" does not have any listed subsectors.

Annual Survey on Philippine Business and Industry (ASPBI)

for the sectors and subsectors
builds on the PSIC (industry classifcation)

Purpose: Provides info about the structure and performance of the Philippines' formal economy
Use: Industry and labor market-oriented reference for localized subsector tagging, from updated

2019 PSIC codes

Context: in 2019, the PSIC was updated from the 2009 version to capture more localized and country specific economic activities. Businesses applying for a business permit in their local governments must register and get tagged under the updated PSIC framework. The 2023 ASPBI reports on 18 of the 21 sectors, with 225 subsectors listed with disaggregated industry data.

Available Data per Subsector:

- Total number of establishments
- Total employment
- Average number of workers per establishment
- Total compensation (in thousand pesos)
- Average annual compensation per paid employee (in pesos)
- Total revenue (in revenue)
- Total expense (in thousand pesos)
- Revenue per expense ratio
- Sales from e-commerce transactions (in thousand pesos)

Republic of the Philippines Professional Regulation Commission (PRC) Boards

just for the HEI-PRC

The Professional Regulation Commission (PRC) of the Republic of the Philippines is the government agency responsible for regulating and licensing various professions in the country. It oversees **Professional Regulatory Boards (PRBs)**, each of which is composed of experts in a specific field (e.g., nursing, engineering, accountancy) who develop and enforce professional standards, administer licensure examinations, and issue licenses to qualified practitioners. Job titles that require passing a PRC licensure exam—such as "Registered Nurse," "Licensed Professional Teacher," or "Certified Public Accountant"—are considered official **Higher Education Institution (HEI) PRC-related job titles**. These titles are legally protected and recognized nationwide, meaning only individuals who hold a valid PRC license for that profession can use them or practice in those regulated fields.

The professions with a requirement for a bachelor's level higher education degree for a board exam taker include:

- Accountancy
- Aeronautical Engineering
- Agricultural Engineering
- Agriculture
- Architecture
- Chemical Engineering
- Chemistry
- Civil Engineering

- Criminology
- Customs Brokerage
- Dentistry
- Electrical Engineering
- Electronics and Communications Engineering
- Fisheries Technology
- Food Technology
- Forestry
- Geodetic Engineering
- Geology
- Interior Design
- Landscape Architecture
- Library Science
- Mechanical Engineering
- Medical Technology
- Metallurgical Engineering
- Midwifery
- Mining Engineering
- Naval Architect & Marine Engineer
- Nursing
- Nutrition and Dietetics
- Occupational Therapy
- Optometrist
- Pharmacy
- Physical Therapy
- Professional Teaching
- Psychology
- Psychometrician
- Radiologic Technology and X-ray Technology
- Real Estate Services
- Respiratory Therapy
- Sanitary Engineering
- Social Work
- Speech-Language Pathology
- Veterinary Medicine

Beyond this list are other professions that are PRC licensed but whose requirements are not bachelor's level degrees (e.g. psychologist, guidance counselor, physician, requiring master's level degrees). To enroll in a program to study law to become an attorney, a learner must first have a

law is quirky like that

bachelor's level degree. Law has its own licensure examination, but administered by the Supreme Court of the Philippines and not PRC.

PSF = Philippine Skills Frameworks PSF is to ensure industry releavance in ICT and Human Capital???

Philippine Skills Frameworks for the ICT Industry (DICT)

The **Philippine Skills Framework (PSF) for the ICT Industry**, spearheaded by the Department of Information and Communications Technology (DICT) through its ICT Industry Development Bureau, defines standardized career maps, occupational roles, skills and competency levels across key ICT sectors—such as **Software Development & Security**, **Analytics & Artificial Intelligence**, **Contact Center & BPM**, **Global In-House Centers**, and **Healthcare Information Management Services**. Officially launched on **May 7, 2024**, the PSF provides a structured roadmap for academic institutions, employers, and training providers to align curricula, job descriptions, upskilling programs, and career progression with actual industry needs.

Philippine Skills Framework – Human Capital

The **Philippine Skills Framework – Human Capital (PSF-HC)** is a government-endorsed competency blueprint for HR and people-management work in the Philippines. It lays out career maps and detailed skill maps for core HR roles—e.g., Chief People/Human Resources Officer, HR Business Partner, Labor Relations, HR Operations & Technology, Performance & Rewards, Talent Attraction/Management, Employee Engagement & Experience, and Learning & Organization Development—each with required competencies and proficiency levels. Organizations, HEIs, and training providers use it to align job descriptions, curricula, and upskilling plans with nationally recognized HR standards and career pathways, as part of the broader Philippine Skills Framework initiative launched by government agencies.

TESDA Workplace Skills and Satisfaction Survey Reports (2020-2024)

TESDA's **Workplace Skills and Satisfaction (WSS) Survey** is a sector-by-sector study used for **Skills Needs Anticipation**—it maps hard-to-fill jobs, emerging skills (incl. 4IR/green skills), training needs, and employer satisfaction with TVET graduates to guide curricula, qualifications, and programs. From **2020–2024**, TESDA released WSS reports by year and sector: **2020—Construction, IT-BPM**; **2021—Logistics, Health**; **2022—Agriculture**; **2023—Manufacturing, Renewable Energy**; **2024—Tourism, Creative**. Each report details sampling, methods (based on an ILO-informed establishment skills survey), skills gaps/shortages, and policy/training recommendations. Key examples include the full reports for **Construction (2020)**, **Logistics (2021)**, **Manufacturing (2023)**, and **Renewable Energy (2023/fielded 2023–2024)**.

TESDA Sector Labor Market Information 2022

TESDA Sector Labor Market Information (LMI) 2022 refers to TESDA's sector-specific reports that analyze current and future skills needs, workforce demand–supply, hard-to-fill occupations, and policy/training implications to guide TVET programs and Training Regulations. A 2022 example is **"Future-Proofing the Construction Sector,"** which profiles the sector's contribution to jobs and output, details emerging skills, and recommends upskilling pathways—material used by TESDA to align curricula and qualifications with industry requirements. The LMI series is curated by TESDA's Planning Office–Labor Market Information Division and published on TESDA's LMI page.

National Workforce Development Plan (led by EDCOM 2, with CHED, TESDA, DepEd)

The **National Education and Workforce Development Plan (NatPlan)** is being led by **EDCOM 2** in partnership with **DepEd, CHED, and TESDA** (with inputs from NEDA, DOLE, DTI and the Private Sector Advisory Council) to align schooling, training, and jobs around priority industries and close skills mismatches. President Marcos publicly backed the NatPlan in January 2025; Congress has since urged its coordination and system-wide implementation. The Senate highlighted a cabinet-level mechanism to oversee it. EDCOM 2 targets delivery of a comprehensive plan to Congress by **late 2025**, defining reforms, budgets, targets, and data systems to synchronize the three education subsectors with workforce needs.

Outputs for Job List Generation: Major Steps

From the inputs, the job titles as outputs were generated across the educational segments in 3 major passes. At each pass, the number of total jobs was adjusted based on steps taken to ensure the list became useable, accurate, and validated.

Pass	Number of Job Titles
• 1st pass of job titles – HEI jobs generated from PSCED and PSOC and combined with TVET and HS	1148
• 2nd pass of job titles – HEI, TVET, and HS jobs together deduplicated	930
• 3rd pass of job titles – Combined job list sector tagged and added with titles from sector reports (industry validation)	1155

Job List Generation: Pass 1 – HEI Jobs Generated from PSCED, PSOC, PRC & TESDA NC HS Jobs

HEI Degree (PRC AND Non-PRC) Level Jobs

Through this segmentation, the 2017 updates to the PSCED was used to identify the bachelor's level degree programs in the Philippines. After extracting a table that contains detailed information about educational clusters, the PSCED qualifications were used to identify likely job titles related to each.

To link qualifications to titles, the 2022 updates to the Philippine Standard for Occupational Codes (PSOC) was used as an additional input. Because the PSOC's index of occupational titles and clusters are meant to represent different possible jobs in the country, it was used as a reference.

With both the PSCED degrees and the PSOC index, a large language model (LLM) was used to generate entry-level job titles for each qualification. To accomplish this, degree programs were pasted into the LLM in batches based on each PSCED broad field cluster. The LLM was prompted to connect each title to a 4-digit PSOC unit code and sample occupational titles from the PSOC index: these codes group related jobs under PSOC major and minor groups.

Inputs	Process	Output
PSOC index of occupational titles PSCED bachelor level degree programs LLM Prompt	Sample prompt chain: https://chatgpt.com/share/688e22c2-c210-8004-b0b7-c6c8b6b71459 2nd prompt: Given this initial list of job titles, diversify the possible job titles associated with the bachelor's level degree and PSOC unit codes to ensure broad representation of possible jobs that are anchored in the Philippine labor market context.	793 titles tagged to HEI degrees

The matched roles were expanded and refined to create diversified job titles while keeping each title tagged to a PSCED qualification and PSOC code.

To confirm HEI titles tagged to PRC titles, each entry tagged as HEI-PRC underwent a verification process that determined their associated PRC license.

Steps:

1. Extracted and isolated jobs that require PRC licensure from the dataset.
2. Determined the associated educational requirement for each PRC license.
3. Determined the appropriate PRC license for each job

Step 1: Extracted and isolated jobs that require PRC licensure from the dataset.

1.1. Using the job list, filtered for jobs tagged under HEI-PRC
1.2. For each entry, extracted the job title and educational requirement

Step 2: Determined the associated educational requirement for each PRC license.

1.3. Used the PRC List of Requirements to map an educational requirement to each license

Input	Process	Output
https://www.prc.gov.ph/uploaded/documents/mListofRequirements.pdf https://www.prc.gov.ph/list-of-requirements	Used an LLM to map each license with an educational requirement Prompt: Given the list of requirements, map each PRC license with the associated educational requirement Manual verification done	PRC educational requirement for each License

Step 3: Determined the appropriate PRC license for each job

1.4. Determined PRC License for each Job Title

Input	Process	Output
Job Titles List of PRC Licenses	Used an LLM to infer the PRC license required based on job title Prompt: Given the following list of job titles, what is the most appropriate license	PRC License for each Job

	required for this role under the Professional Regulation Commission (PRC) of the Philippines? Manual verification done	

1.5. Determined PRC License for each Educational Requirement

Input	Process	Output
Educational Requirements List of associated degrees for each PRC License	Used an LLM to infer the PRC license required based on educational requirement Prompt: Given the following list of educational requirements, what is the most appropriate license under the PRC? Manual verification done	PRC License for each educational requirement

1.6. Compared the outputs of 3.1 and 3.2. If the outputs match, wrote the PRC license of the entry into the main job list. If not, manually verified what the associated PRC license is for the entry.

Summary of extracted PRC professions and their HEI associated degree requirements

Profession	PRC Degree Requirement
Accountancy	B.S. Accountancy
Aeronautical Engineering	B.S. in Aeronautical Engineering or equivalent
Agricultural and Biosystems Engineering	B.S. in Agricultural Engineering or equivalent
Agriculture	Bachelor's degree in Agriculture (General Course) or B.S. in Agriculture with specific majors
Architecture	B.S. in Architecture
Chemical Engineering	B.S. in Chemical Engineering or equivalent
Chemistry	B.S. in Chemistry or Allied Degree under CHED Technical Committee for Chemistry
Civil Engineering	B.S. in Civil Engineering
Criminology	B.S. in Criminology
Customs Brokers	B.S. in Customs Administration
Dentistry	Doctor of Dental Medicine or equivalent

Electronics Engineering	B.S. in Electronics & Communication Engineering or B.S. Electronics Engineering or equivalent/related engineering course
Fisheries	Bachelor of Science degree in any field of Fisheries
Foresters	Bachelor's degree in Forestry
Geodetic Engineering	B.S. in Geodetic Engineering
Geology	B.S. in Geology
Interior Design	B.S. in Interior Design
Landscape Architecture	B.S. or post-graduate degree in Landscape Architecture
Librarianship	Bachelor of Library and Information Science
Master Plumbing	Bachelor's degree in Architecture/ME/CE/Chemical Engineering/Mining Engineering, or High School Diploma/Transcript (Notarized) + 5 years of experience
Mechanical Engineering	B.S. in Mechanical Engineering
Medical Technology	B.S. in Medical Technology, B.S Public Health
Metallurgical Engineering	B.S. in Metallurgical Engineering or related engineering degree majoring in metallurgical engineering/metallurgy
Midwifery	Graduate Midwife or registered PRC Nurse with additional training
Mining Engineering	B.S. in Mining Engineering
Naval Architecture	B.S. in Naval Architecture and Marine Engineering
Nursing	B.S. in Nursing
Nutrition and Dietetics	B.S. in Nutrition and Dietetics
Occupational Therapy	B.S. in Occupational Therapy
Optometry	Doctor of Optometry (six-year course)
Pharmacy	B.S. in Pharmacy or equivalent
Physical Therapy	B.S. in Physical Therapy
Psychometrician	Bachelor's degree in psychology from CHED recognized institution
Radiologic Technology	B.S. in Radiologic Technology
Real Estate Service	Bachelor's degree in Real Estate Services, Relevant bachelor's degree from state university or CHED-recognized institution
Real Estate Service	Bachelor's degree in Real Estate Services, Relevant bachelor's degree from state university or CHED-recognized institution
Electrical Engineering	B.S. in Electrical Engineering
Respiratory Therapy	B.S. in Respiratory Therapy from CHED recognized institution
Sanitary Engineering	B.S. in Sanitary Engineering or BS in Civil Engineering with major subjects in sanitary engineering

Social Workers	B.S. in Social Work
Teachers	Bachelor's in Elementary Education or equivalent; Bachelor's degree in Early Childhood Education (BECED) or equivalent; Bachelor's degree in education or equivalent with major and minor
Veterinary Medicine	Doctor of Veterinary Medicine from CHED accredited College
Food Technology	Bachelor of Science in Food Technology or its equivalent (BS Food Science and Technology, BS Food Engineering, BS Food Technology and Entrepreneurship, BS Food Science, BS Food Science and Nutrition, BS Nutrition and Food Technology, and Bachelor of Food Technology)
Speech-Language Pathology	Bachelor of Science degree in Speech-Language Pathology
Optometry	Doctor of Optometry

TESDA NC Level Jobs

Through contact with the TESDA Competency Standards Development Division – Qualifications and Standards Office (TESDA QSO), a primary source was obtained. It derives directly from a TESDA database. TESDA QSO shared two job list files, one with tagging to training regulation (TR) programs and the other for competency standard (CS) programs.

A **Training Regulation (TR)** is TESDA's official document that specifies the competency standards, assessment arrangements, and curriculum for a particular TVET qualification—essentially the blueprint for delivering and assessing a complete training program that leads to a national certificate (NC).

A **Competency Standard (CS)**, on the other hand, defines the skills, knowledge, and attitudes required to perform a specific job or function, but it is not yet packaged into a full training program. CS documents are often used as building blocks for developing TRs, or for workplace-based training and assessment where a full qualification is not needed.

The TR file had 340 original rows, while the CS file 169 rows (combined n = 509). Through this tagging, the list of TESDA NC level jobs came directly from the national government agency responsible for technical-vocational skilling.

Filtering Process for Updated Jobs List:

- The TR and CS files were combined into one sheet
- For the first round of filtering (n = 392), additional columns were added to tag rows based on: if TR or CS, if aspirational job, if entry-level, and if NC Level I-IV

- 10 jobs were tagged as non-aspirational, and therefore excluded from the final list (9 being NC Level I)
- 67 jobs were tagged as not entry-level based on filtering for keywords, and manual verification by checking online (NC Levels III-IV jobs)
- 22 rows were tagged as not a job, wherein the qualification offered by TESDA does not lead to an easily identifiable job

- After filtering for aspirational and entry-level jobs, the list of jobs was manually inspected to suggest job titles per row, and to deduplicate cases where the same or similar jobs were present in different rows (assigned to different qualifications)
- After the second round of filtering (n = 346), the updated list of jobs can be found in the FILTERED_Combined_2nd pass tab at Column H

Schema for updated jobs list table (346 rows):

Column	Description
Sector	The sector tagging TESDA used for the qualification and related job. TESDA has its own sector tagging of qualifications that is different from industry tagging done by the Philippine Statistics Authority in the Labor Force Survey (LFS).
Qualification Title	The name of the TESDA qualification offered at a training institute.
Level	Manual tagging done (not by TESDA) to evaluate between aspirational and non-aspirational jobs.
Job Titles (From TESDA)	The original list of job titles, delimited by “,”, in the source files shared by TESDA QSO, per qualification row. In rows where job titles were merged across similar qualifications, the list of original titles are pasted into one cell.
TR or CS	TR refers to Training Regulation while CS refers to Competency Standard. TESDA programs with TRs let graduates obtain National Certificates (NCs) that must be validated through assessments; Philippine employers seek assessed graduates. TESDA programs with only CS do not offer National Certificates. For the purpose of TESDA’s enrollment and graduation tracking, learners tagged under CS are tracked separately. Currently, the foundation team has access to graduates of TR certified programs from 2024.

Entry Level	Manual tagging done (not by TESDA) to check if the indicated role for the qualification is open to entry-level candidates. Keywords such as Manager, Lead, Executive, etc. were used to filter for non-entry-level jobs.
NC Level	NC refers to National Certificate. NC Level refers to the skills complexity has tagged specific programs with. In general, NC Levels I-II are entry-level in nature while levels III-IV are for more advanced skills. In the job list, there are roles tagged levels III-IV but still considered entry-level: these are technician and skilled worker jobs. They require more technical knowledge but do not necessarily require more years of experience, or to manage teams.
Suggested Job Titles	Suggested names of jobs per row that were manually confirmed from the list of job titles TESDA indicated per qualification. In cases where specific qualifications taught different job tasks done by a principal role, those rows were merged back into one job. This is also true for job rows TESDA separated by skill level (e.g. Plumber I, Plumber II, Plumber III = Plumber). Conversely, in cases when the TESDA indicated list of job titles referred to mutually distinct functions, those titles were unpivoted into separate rows.

Senior High School Diploma Level Jobs

Two approaches were done to generate the high-school level jobs: 1) scraping job posts from Philippine jobsites that only require high school diplomas, and 2) reviewing studies that discuss employer preferences for the kinds of jobs senior high school graduates are considered for.

Step-by-Step Calculation Process:

Step 1: Data Sourcing and Initial Assessment

Data Source: The raw dataset consisted of a total of 1,258 job postings, assembled through Octoparse implementation, a web-based, visual scraping tool.

Initial Analysis: The raw dataset comprised 1,258 entries, including 42 unique but highly inconsistent job titles. A preliminary assessment revealed that many titles were convoluted with non-essential information (e.g., "URGENT HIRING | Non Voice Live Chat Support | Fresh Grads are welcome to apply!"). The data was loaded into a pandas DataFrame for analysis.

Step 2: Noise Assessment and Filtering

Rationale: Given that raw job postings contain excessive promotional language, location-specific details, salary information, and company branding, a noise filtering step was essential to enhance data quality and usability. A quantitative scoring system was therefore designed to improve the signal-to-noise ratio for subsequent natural language processing (NLP) tasks by determining and removing job postings with excessive noise.

Technical Implementation: A rule-based scoring system was designed in Python to evaluate and filter out noisy job postings using a two-tier noise detection approach:

- **Noise Detection Framework**
    - **Minor Noise:** captured via a **noise_patterns** dictionary, which included:
        - urgency_keywords (e.g., "URGENT HIRING", "URGENT")
        - experience_tags (e.g., "Fresh Grads are welcome to apply", "Fresh Grads can JOIN")
        - salary_benefits (e.g., "Up to 24K Salary", "50K Monthly Incentives")
        - locations (e.g., "BGC", "Shaw, Mandaluyong")
        - work_arrangements (e.g., "Virtual Process Only", "WFH")
        - company_names (e.g., "CITI Phone Professional", "ZARA")
        - job_posting_fluff (e.g., "PROJECT BASED", "is a plus")
        - employment_types (e.g., "Full-Time", "Part-Time")
    - **Severe Noise:** detected by employing **exclusion_patterns,** a list of regular expressions (regex) specifically engineered to match overly promotional or non-standard job postings (e.g., full job ads embedded in title fields).
- **Scoring and Filtering System:** To operationalize the filtering process, a custom scoring function was developed to assign a noise score to each job posting. Each posting's score is initiated at 0 and updated based on the following system:
    - **Scoring Logic:**
        - **+3 points** for each match with an exclusion pattern
        - **+1 point** for each match with a noise pattern
    - **Filtering Rule:** A noise threshold of 2 was defined:
        - Postings scoring $>$ 2 were classified as "excessively noisy" and removed, considering that these are highly likely to be non-job descriptions.
        - Postings scoring $\leq$ 2 were retained for further processing.

Outcome: Out of the 1258 web-scraped job postings, **507 entries were removed**. The resulting dataset retained **751 postings,** encompassing **29 unique titles**, thus realizing a **data retention rate of 59.7%** and a substantially cleaner input for subsequent analysis.

Input	Process	Output

Raw Job Postings (N=1,258) Unique Job Titles (N=42)	Apply the assess_noise_level function: • +1 for noise_pattern matches • +3 for exclusion_pattern matches Apply a threshold of 2	Filtered Postings (N=751) Unique Job Titles (N=29)
Example: Restaurant Staff (Server, Cashier, Line Cook, Dishwasher, Receptionist)	Score: 1 → Retained	Restaurant Staff (Server, Cashier, Line Cook, Dishwasher, Receptionist)
Example: URGENT HIRING Non-Voice Live Chat Support	Score: 3 → Excluded	

Table 1. Input–Process–Output (IPO) Summary: Noise Filtering

Step 3: Title Cleaning and Regex-Based Standardization

Rationale: Although the previous noise filtering step successfully removed irrelevant postings, the retained titles still exhibited superficial textual variations that could mislead downstream semantic analysis. For example, "CSR," "customer service rep," and "customer service representative" all refer to the same job but would be treated as three (3) distinct entities by a simple text-matching algorithm. To ensure the semantic model accurately groups titles based on meaning rather than on minor spelling or formatting differences, the text needed to be normalized. This cleaning process functions as a crucial prerequisite for credible semantic grouping.

Technical Implementation: The **advanced_clean_title** function was designed to systematically normalize the text of each job title using a series of re.sub (regex substitution) operations. These operations included:

- **Case Folding:** Converting all text to lowercase to ensure, for example, "Server" and "server" are treated identically.
- **Character Removal:** Stripping special characters and symbols such as |, -, and () that might have remained in the titles.
- **Noise Term Stripping:** Removing the minor noise terms identified in the previous step (e.g., location names) to ensure they do not interfere with the core title's meaning.
- **Abbreviation Standardization:** Another dictionary of regex patterns was used to expand and unify common abbreviations. For instance:
    - **Pattern: r'\\b(csr|customer service rep)\\b'**

- **Replacement: “customer service representative.”** This ensured that all equivalent abbreviations were mapped to a consistent canonical form.

Outcome: While the number of total job posting entries **remained at 751**, this normalization process trimmed the number of unique job titles **from 29 to 25**. This reduction confirms that semantically equivalent titles were correctly consolidated, implying successful standardization and data preparation for deeper semantic clustering in the next step.

Input	Process	Output
Filtered Postings (N=751) Unique Job Titles (N=29)	Apply advanced_clean_title() function: • Normalize case (lowercase) • Remove special characters and delimiters • Strip noise terms (e.g., locations, salaries, fluff) • Standardize abbreviations via regex • Extract multiple roles from parentheses Steps often apply in combination to the same title.	Cleaned Postings (N=751) Cleaned Unique Job Titles (N=25)
Example: Cashier, CASHIER	Case Folding (Lowercasing)	cashier
Example: Restaurant Staff (Server, Cashier, Line Cook, Dishwasher, Receptionist)	Role splitting + Character Removal	restaurant staff server, cashier, line cook, dishwasher, receptionist
Example: BACK OFFICE - CSR | START ASAP! | Shaw, Mandaluyong	Noise Term Stripping + Abbreviation Standardization	back office customer service representative
Example: csr, CSR	Abbreviation Standardization	Customer Service Representative

Table 2. Input–Process–Output (IPO) Summary: Title Cleaning

Step 4: Semantic Clustering

Rationale: Despite being reduced to 25 unique job titles after cleaning and standardization, many titles in the dataset remained semantically similar (e.g., “Factory Worker” vs. “Production Worker”). Simply put, reliance on conventional keyword-based methods (e.g., string matching or token frequency) would prove inadequate in capturing these foundational similarities or would fail to acknowledge the equivalence between semantically similar but lexically different phrases. As a result, a pre-trained language model was harnessed to objectively quantify the degree of similarity between titles for a more robust and validated basis for clustering.

Technical Implementation:

- **Embedding Generation:** The first stage is to convert each unique job title from human-readable text into a machine-readable, numerical format, particularly vectors. This was accomplished using the **SentenceTransformer(‘paraphrase-MiniLM-L3-v2’)** model, a lightweight pre-trained deep learning model known for its power and ability to apprehend contextual meaning beyond mere “keywords.” This is through generating high-dimensional vector embeddings representative of the semantic content of sentences. Still, before embedding, a deduplication step was performed using **.unique()** to double check and extract only the distinct job titles, optimizing performance by avoiding redundant computations (e.g., processing "Customer Service Rep" only once regardless of its frequency). These unique titles were encoded in batches via **model.encode(batch, ...)**, generating embeddings such that semantically similar titles (e.g., "Factory Worker" and "Production Worker") are numerically close in vector space.
- **Similarity Matrix:** With embeddings in place, a similarity matrix was constructed using **cosine_similarity** from **sklearn.metrics.pairwise**. This matrix quantifies how semantically similar each pair of titles is:
    - **1.0:** Identical semantic direction
    - **0.0:** No relation
    - **-1.0:** Opposite meanings
    - The result is a square matrix where the (i, j) entry returns the similarity between title i and j (i.e., a complete semantic relationship map).
- **Hierarchical Clustering: AgglomerativeClustering** was then applied using the similarity matrix. This **“bottom-up”** approach starts with or treats each title as its own cluster and merges the most similar pairs iteratively until a threshold is breached.
    - **Key Parameters:**
        - **n_clusters=None:** The algorithm does not fix the number of clusters but stops when the distance exceeds the threshold.
        - **distance_threshold=1 - threshold:** The clustering algorithm works with distance (where a low value means high similarity), but a **similarity matrix** (where a high value means high similarity) is also present. As such, the similarity scores (e.g., **0.95**) are converted into distance scores (1 - 0.95 = 0.05). The algorithm will only merge clusters whose distance is less than or equal to this threshold.
        - **linkage=‘complete’:** Defines intercluster distance as the farthest pairwise distance between points in two (2) clusters, which encourages tight, compact clusters.
        - **metric=‘precomputed’:** This is an efficiency setting. It instructs the algorithm to use the **1 - similarity_matrix** that was already calculated, rather than re-computing distances.

- **Exploration at Multiple Thresholds:** Exploration was performed at three semantic thresholds (0.95, 0.85, 0.75) to evaluate the sensitivity of the groupings:
    - **0.75 (Permissive):** Produced 21 clusters by consolidating broader groups, sometimes over-merging distinct titles.
    - **0.85 (Balanced):** Formed 24 clusters by merging closely related roles.
    - **0.95 (High Precision):** Created 25 clusters, retained everything.
        - **Justification:**
            - **Integrity of Role Classification:** The salient reasoning behind such a choice was the preference for precision over recall. In other words, "Type I errors" or semantic misgroupings (e.g., incorrectly merging "Construction Laborer" and "Maintenance Laborer") were wanted to be prevented.
            - **Clean Foundation for Manual Review:** Considering the previous reasoning and the choice, it is palpable that the chief objective was not final categorization but a precise starting point. This means that domain experts or human critics would later join the loop, review, and merge clusters (e.g., "Customer Service" vs. "Customer Service Rep") with thorough semantic traceability.
            - **Principal Component Analysis (PCA):** Clustering manifests in a high-dimensional space. However, PCA has been utilized to project embeddings into 2D space for visual inspection. PCA has its advantages, but it tends to oversimplify. Specifically:
                - On PCA, embeddings may appear to be clustered and extremely close, but in reality, they may differ meaningfully in original dimensions.
                - At a 0.95 threshold, PCA may display embeddings that serve similar roles in close proximity, but the model properly distinguishes them because of minute semantic distinctions.

Outcome: The final clustering produced semantically tight and high-integrity groups: 0.95 threshold (preserved all 25 job titles as distinct clusters), 0.85 threshold (reduced to 24 clusters), 0.75 threshold (merged further to 21 clusters, albeit with reduced precision). As observed, this multi-threshold setup cements and puts human judgment at the forefront of the process, notably the trade-offs between precision and coverage, leading to a defensible and flexible clustering foundation.

Input	Process	Output
Cleaned Unique Job Titles (N=25)	Using the **SentenceTransformer** model to generate embeddings, a **cosine_similarity** matrix was then computed, and **AgglomerativeClustering** was	Algorithmically-Derived Clusters: • 0.95 Threshold (N=25) • 0.85 Threshold (N=24) • 0.75 Threshold (N=21)

	applied to group the titles according to their semantic similarity.	
Example: server, line cook, dishwasher, cashier, receptionist, restaurant staff		**Cluster X:** {server, line cook, dishwasher, cashier, receptionist, restaurant staff}

Table 3. Input–Process–Output (IPO) Summary: Semantic Clustering

Step 5: Representative Title Selection

Rationale: Since clusters were formed already, the next task is the selection of each cluster's representative title. However, instead of selecting the title with the highest frequency within each cluster, a multi-step or multi-metric scoring system is again structured to identify one "canonical" title per cluster that is semantically central, statistically significant (i.e., high frequency), and clean and professional. This ensures each selected title can serve as a clear and accurate stand-in for its entire group.

Technical Implementation: A custom function **select_improved_representative()** is executed to compute an aggregate score for every title within a cluster, wherein the title with the highest overall score becomes the representative. The scoring system is a weighted combination of five (5) distinct metrics:

- **Centrality Score:** This metric contains the highest weight and thus the most important one, as it is the technical implementation of "semantic representativeness."
  - **Purpose:** Measures how close each title is to the "semantic center" of its cluster.
  - **Method:** Compute the average embedding vector for the cluster, then calculate the cosine similarity of each title to this average.
  - **Rationale:** Selects the title that is most aligned with the cluster's overall meaning.
- **Frequency Score:** This metric is a direct measure of a title's prevalence in the original dataset.
  - **Purpose:** Demonstrate how often a title appears in the original data.
  - **Method:** Normalize raw counts within the cluster.
  - **Rationale:** Frequently occurring titles are likely to be standardized, professional terms.
- **Length Score:** This is a penalty-based heuristic to favor concise titles.
  - **Purpose:** Favors concise and easily understandable titles.
  - **Method:** Apply a penalty for longer titles (e.g., 1 / [number of words]).
  - **Rationale:** Shorter titles are more usable in reports and visualizations.
- **Professionalism & Redundancy Scores:**
  - **Purpose:** Enforce quality control based on domain knowledge.
  - **Professionalism:** bonus for titles containing professional keywords like "specialist," "officer," "manager," or "technician."
  - **Redundancy:** Penalty for repetitive or filler words.
  - **Rationale:** Prevents informal, low-quality, or poorly-phrased titles from being selected.
- **Weighted Combination:** The five scores are combined into a single, final score using a weighted average. The weights are manually tuned based on which factors are deemed most important. For example:
  - **Centrality Score:** 0.35

- **Frequency Score:** 0.25
- **Length Score:** 0.15
- **Professionalism & Redundancy Scores:** 0.25

Outcome: As a result of this step, the system generated **25 representative titles**, one for each of the clusters generated by the 0.95 threshold. The reported average confidence score of 1.0 is a direct result of the high-precision clustering. Since the 0.95 threshold was exceedingly conservative in grouping only nearly identical titles, the "semantic center" of each cluster was exact for each title in it, resulting in a perfect average cosine similarity score.

Input	Process	Output
Algorithmically-Derived Clusters (N=25)	For each cluster, the **select_improved_representative** function was applied to evaluate and select the single, most representative title based on the following weights: • **Centrality Score:** 0.35 • **Frequency Score:** 0.25 • **Length Score:** 0.15 • **Professionalism & Redundancy Scores:** 0.25	Representative Titles (N=25)
Example: Cluster Y: {Customer Service Representative (CSR), Customer Service Representative (CSR)...}		customer service representative

Table 3. Input–Process–Output (IPO) Summary: Representative Title Selection

Step 6: Final Semantic Grouping ("Umbrella" Categories)

Rationale: This is the culmination of the entire data pipeline. The clustering steps (Steps 4 and 5) were designed for technical precision, creating a large number of tight clusters. While this upholds data integrity, as mentioned previously, some of the titles remain semantically equivalent but lexically different, which is not ideal for reporting. In short, this final grouping stage serves as a "human-in-the-loop" consolidation layer (i.e., a supervised, domain-specific categorization step that reinforces the output of the automated, unsupervised clustering process).

Technical Implementation:

- **The umbrella_groups Dictionary:** This is the core of the entire step. It is a manually created, human-defined Python dictionary that explicitly maps the **representative titles** (the output of Step 5) to the final umbrella category names, wherein the:
    - **Keys:** The final, standardized umbrella category names (e.g., "Customer Service Representative")
    - **Values:** Lists of the representative titles that belong to that category (e.g., ["customer service representative," "customer service officer," "back office staff," # ... other related representative titles])
- **The apply_improved_umbrella_grouping Function:** This function automates the mapping process across the entire dataset. It takes the **df_clean** DataFrame and the **umbrella_groups** dictionary as input.
    - **Process:** The function iterates through each row of the DataFrame. For each row, it checks the **representative title** of the cluster it belongs to. It then searches for this title within the **umbrella_groups** dictionary.
    - **Matching Logic:** The function uses a combination of matching strategies:
        - **Exact String Matching:** It first tries to find an exact match for the representative title in the dictionary's lists.
        - **Partial String Matching:** For added validity, it can also use partial string matching (e.g., matching "admin staff" to "administrative staff") to catch minor discrepancies.
    - **Output:** A new column, **umbrella_group**, is created in the **df_clean** DataFrame. Each row is assigned the standardized umbrella category name that corresponds to its representative title.

Outcome: Out of the 25 representative titles, 16 final highly accurate and standardized job categories remained, rendering a reduction ratio of **2.6x** from the initial unique raw titles.

Rank	Job Title	Frequency	Percentage (%)	Cumulative (%)	Category
1	Customer Service Representative	195	25.97	25.97	High Frequency
2	Cashier	195	25.97	51.94	High Frequency
3	Housekeeper	78	10.39	62.33	Medium Frequency
4	Restaurant Staff	78	10.39	72.72	Medium Frequency
5	Production Worker	40	5.33	78.05	Medium Frequency

6	Warehouse Associate	40	5.33	83.38	Medium Frequency
7	Store Crew	39	5.19	88.57	Medium Frequency
8	Pharmacy Associate	39	5.19	93.76	Medium Frequency
9	Call Centre Staff	39	5.19	98.95	Medium Frequency
10	Delivery Driver	2	0.27	99.22	Low Frequency
11	Masseur	1	0.13	99.35	Low Frequency
12	Maintenance Worker	1	0.13	99.48	Low Frequency
13	Painter	1	0.13	99.61	Low Frequency
14	General Laborer	1	0.13	99.74	Low Frequency
15	Heavy Equipment Operator	1	0.13	99.87	Low Frequency
16	Sales Representative	1	0.13	100.00	Low Frequency

Table 4. Final Job List

Input	Process	Output
Representative Titles (N=25)	Each representative title is matched to a standardized umbrella category using a supervised, domain-informed dictionary (**umbrella_groups**). The	Final Job Titles (N=16)

	apply_improved_umbrella_grouping() function automates this with: • **Exact match** to dictionary entries • **Partial match** for close lexical variants	
Example: • Cluster Y: customer service representative • Cluster Z: back office customer service representative • Cluster B: customer service specialist	**umbrella_groups** = Customer Service Representative: [“customer service representative,” “customer service specialist,” “back office customer service representative,” “collection transfer agent”...] Cluster Y, Z, B → Exact Match → Customer Service Representative	Customer Service Representative
Example: • Cluster A: customer service	Cluster A → No Exact Match → Proceed to fuzzy/partial matching “customer service” matches significantly with umbrella keywords (If ≥ 50% of variation words are in the title, it is a match) → Assign umbrella category	Customer Service Representative

Table 5. Input–Process–Output (IPO) Summary: Final Semantic Grouping

Python Notebook Link: [Final HS Validation (For Review).ipynb]

Recombination into Single Data Schema

Job titles across different education segments were recombined into one table with the following schema:

Field Name	Description
Segment (Education)	HEI or non-HEI delineation
Job Title	Distinct job title
PSOC Unit Code	4-digit unit code from 2022 PSOC matched

Alternative Title	Sample PSOC unit code title
Degree programs	Minimum educational qualification
Source	Indicated data source

Pass 1 Summary

Inputs	Outputs	Value
PSCED + PSOC + LLM Prompt	HEI – Non-PRC	616
PSCED + PSOC + LLM Prompt	HEI - PRC	177
Webscrape – HS Jobs	HS	16
TESDA QSO - TESDA jobs	TESDA NC	339
Total		**1148**

Job List Generation: Pass 2 – HEI, TVET, and HS jobs together deduplicated

The deduplication process began with the Pass 1 list of jobs, removing or merging duplicates for greater accuracy. Exact matches were recombined into single rows. For near-duplicates, similar job rows were clustered based on lexical similarity and reviewed using two parameters: similarity of job titles and similarity of associated degree programs. These clusters were then examined to merge synonymous or closely related titles, with a special check for HEI PRC degrees to ensure updated titles were correctly inserted. Jobs that were in multiple education segments were recombined into the lowest minimum qualification. The final output was a cleaned Pass 2 job list with duplicates removed or recombined.

Sector tagging was done for the job title rows, basing from the PSIC and ASPBI economic sectors.

Input	Process	Output
- 930 deduplicated job titles - PSIC / ASPBI list of (19) sectors	Used an LLM to help tag jobs back to economic sectors Prompt: Based on what you can infer about these job titles’ responsibilities, the competencies needed for each job (also basing from their qualifications), and the labor market context they are likely situated in, which economic sector tagging is best for each?	930 deduplicated job titles tagged to an economic sector

	[Paste 10-20, 30-50 jobs at a time] Manual verification by checking online	

In order to tag jobs to sectors, an LLM prompt was used to determine the likely tagging. To ensure accuracy and validity, the prompt instructed an LLM to assess the competencies of a job role and the likely field a job is situated in. This process was done first to tag jobs to sectors, before the 3rd pass of the method wherein job titles from sector reports were added back to the list. In the 3rd pass, sector tagged jobs were double checked to correct mistakes, and to add in subsectors as the next level of tagging.

Summary of Job Titles by Sectors in the 2nd pass

Sectors	Job Titles (Pass 2 – deduplicated, sector tagged)
Manufacturing	141
Professional, Scientific and Technical Activities	141
Human Health and Social Work Activities	92
Information and Communication	92
Agriculture, Forestry, and Fishing	75
Construction	61
Education	57
Arts, entertainment and recreation	47
Accommodation and Food Service Activities	36
Administrative and Support Service Activities Sector	34
Public Administration and Defense; Compulsory Social Security	31
Transportation and Storage	31
Financial and Insurance Activities	30
Electricity, Gas, Steam and Air Conditioning Supply	21
Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles	14
Other Service Activities	12
Water Supply; Sewerage, Waste Management and Remediation	7
Fishing and aquaculture	3
Real Estate Activities	3
Total	**930**

Pass 2 Summary

Inputs	Old Value	New Value
HEI - Non PRC	616	463
HEI – PRC	177	137
HEI – PRC (Post-Bachelor's)	X	5
HS Jobs	16	16
TESDA jobs	339	309
Grand Total		**930**

Note: HEI-PRC jobs were delineated between those that only require a bachelor's level degree versus specific professions that require a master's level degree. Those jobs are excluded from the final list of entry-level jobs to be included in the MCA tool.

Job List Generation: Pass 3 – Industry validation through sector reports

The **third pass** of the job list compilation focused on enriching the dataset by integrating job titles drawn from multiple sector-specific reports, ensuring that the list reflected both breadth and depth across industries. Using the **Pass 2 list of jobs recorded by sector** as the starting point, job titles from **Philippine Skills Framework (PSF) reports**, **TESDA Workplace Skills and Satisfaction Survey (WSSS) reports**, and **TESDA Labor Market Information (LMI) reports** were reviewed and added into their corresponding sectors.

Sector Report Source	Industry	Job Titles
Philippine Skills Framework	Analytics and Artificial Intelligence	
Philippine Skills Framework	Contact Center and Business Process Management	
Philippine Skills Framework	Human Capital	
TESDA Workplace Skills Survey Report (WSSR)	IT-BPM	
TESDA Workplace Skills Survey Report (WSSR)	Manufacturing	
TESDA Workplace Skills Survey Report (WSSR)	Tourism	
TESDA Workplace Skills Survey Report (WSSR)	Agriculture	

TESDA LMI Report	Construction	
TESDA LMI Report	Creative Industry	
TESDA LMI Report	Health and Wellness	
TESDA LMI Report	Transportation and Logistics	

For **sector-specific jobs**, only those titles that were not already in the sector list were added to avoid duplication. This ensured that new roles from targeted industries such as Analytics and Artificial Intelligence, Contact Center and Business Process Management, Human Capital, IT-BPM, Manufacturing, Tourism, Agriculture, Construction, Creative Industries, Health and Wellness, and Transportation and Logistics expanded the coverage without inflating overlaps.

For **sector-agnostic jobs**—roles that could exist across multiple industries or were not tied to one specific sector—many were identified in the source reports. Some of these were reintroduced into the dataset, especially if they had been omitted in earlier passes, to preserve comprehensiveness in the overall job inventory.

After integrating these titles, the entire dataset underwent a **sector tagging recheck** to correct any misplacements, ensuring that each role was aligned with its most relevant sector. Finally, the updated sector-tagged job titles were assigned **ASPBI subsectors** to provide consistent economic classification, resulting in **Pass 3**—a more complete, sector-enriched, and economically categorized job list.

Pass 3 Summary

Inputs	Old Value	New Value
HEI Non-PRC Jobs + Sector Reports	463	583
HEI PRC Jobs Jobs + Sector Reports	137	163
HEI PRC (Post-Bachelor's)	5	6
HS Jobs + Sector Reports	16	48
TESDA jobs + Sector Reports	309	361
Grand Total		**1155**

Validation

The summary for generating the pass of jobs is as follows:

- Generated job titles from degree programs and PSOC codes, TESDA qualifications, and online scraping of high school–level roles.
- Tagged each job by sector and subsector using ASPBI macroeconomic data
- Deduplicated/reduced jobs based on lexical similarity, via clustering and content review.
- Reviewed job list and enriched with industry input by checking sector skill council reports from Philippine Skills Framework and TESDA

This brought us to a cleaned list of approximately 1,155 jobs with sector assignments.

Partner Validation

As part of the validation process, the MCA team engaged with EDCOM 2 (the Second Congressional Commission on Education), a key government stakeholder. We shared the methodology behind the development of the 1,100-job list—including how job titles were generated, tagged to sectors and subsectors, deduplicated, and cross-referenced with ASPBI economic data for validation. In response, EDCOM 2 provided valuable insights drawn from their own industry consultations and sectoral dialogues. Their feedback helped us further align the job list with current labor market trends and priority skills needs identified at the national level. This collaboration reinforced the credibility of the job list and ensured that it reflected not only statistical indicators but also qualitative signals from government and industry partners.

Youth Validation

To ensure the MCA tool reflects the real experiences and aspirations of Filipino learners, a round of youth validation was conducted through online focus group discussions (FGDs). These FGDs were not usability tests, but deep conversations designed to surface how young people make career decisions, how they feel about digital guidance tools, and what job roles they see—or don't see—in their schools and communities.

<table>
<tr><th>FGD Group</th><th>Description</th><th>Participants</th></tr>
<tr><td>Senior High School students from Christian high school in Santa Rosa City</td><td rowspan="2">Online focus group discussions were organized between August 9-10</td><td>9</td></tr>
<tr><td>University students from Region 4A universities (public and private)</td><td>10</td></tr>
</table>

University students from Metro Manila (private, top schools)		6
Mixed group of college level students from Quezon City, convened by Positive Youth Development Network (PYDN)	A live poll about main career influences was conducted, at a PYDN event during International Youth Day (August 12)	~20

Online Focus Group Discussion Questions

Category	Purpose
A. Career Agency & Influences	Understand youth context, validate tool relevance
1. When you think about your future career, what kind of decision-making feels within your control? What feels out of your control? 2. Who or what influences your career decisions the most right now? (Family, teachers, social media, job trends, etc.) 3. What tools or resources (if any) have helped you feel more confident about choosing a career path?	
B. Emotional Landscape	Surface motivations and fears MCA must be sensitive to
4. What makes you feel excited or nervous when thinking about your future career?	
C. MCA Tool Framing & Trust	Test perceived usefulness, openness, trust conditions
5. Imagine a tool like My Career Advisor that helps suggest job roles and learning paths. What would you expect it to help you with? 6. Would you be open to using a tool that gives career suggestions based on your answers or interests? Why or why not?	

7. What would make you trust or distrust a tool like this?

Live Poll Questions

Statement	Response
My family's choice is my choice when it comes to my future career	Agree or Disagree
My favorite teacher or my favorite subject influences my career choice	
Doing a job that I love is very important to me	
Doing a job that makes a lot of money is very important to me	

Main Themes & Insights from Youth FGDs

1. Career Decision-Making: Balancing Control and Uncertainty

- Young people feel **agency over personal choices**—such as selecting their course, developing skills, applying to opportunities, and building networks—but recognize limits when it comes to **market demand, economic conditions, and sociopolitical realities**.
- Across contexts, there is tension between **personal aspirations** and **external constraints**, whether these come from the labor market, family expectations, or unpredictable events ("kapalaran").

2. Strong and Complex Role of Family

- Family is a **dominant influence** in career decisions, providing both **support** and **pressure**.
- For some, family is a **safety net** that enables risk-taking and exploration. For others—especially first-generation college-goers—it is a **gatekeeper**, directing choices toward perceived stability and away from certain institutions or fields.
- Family values, life experiences (e.g., OFW work, entrepreneurship, charity work), and economic needs shape both career direction and urgency.

3. Influence of Mentors, Peers, and Role Models

- Teachers, professors, and mentors help students **identify strengths** and expand career possibilities they may not have considered.
- Peers and seniors provide real-world insights about courses and jobs.
- Informal role models—including online creators—fill gaps when family or school guidance is limited.

4. Heavy Reliance on Digital and Social Platforms

- Students turn to **online resources** for career exploration, skill-building, and insider perspectives.
    - **High school students**: Facebook, YouTube podcasts, seminars.
    - **College students**: LinkedIn, Reddit, ChatGPT, Khan Academy, TikTok, Instagram reels.
- Social media algorithms and content creators strongly influence perceptions of career viability, often shaping motivation and confidence.

5. Motivations: Purpose, Growth, and Support for Others

- Many students are motivated by **helping others—**whether through teaching, charity, justice, or community work—and see their careers as a way to give back to family and society.
- Career excitement often comes from the **opportunity to apply skills**, **meet people**, and **grow personally and professionally**.

6. Persistent Career-Related Fears

- Common fears include:
    - **High competition** in saturated labor markets.
    - **Mismatch** between graduate qualifications and employer demands.
    - **Low pay vs. high cost of living**, particularly in urban centers.
    - **Unemployment** and fear of education "going to waste."
    - **Political instability** affecting local and overseas opportunities.
    - **Burnout** from jobs that don't align with expectations or values.

7. Gaps in Career Guidance

- Students want **clearer, more specific pathways—**including step-by-step learning plans, industry-aligned skills lists, and local job market insights.
- They seek help in **translating broad skills into concrete career options**, especially when they have multiple interests but no clear direction.

8. Expectations for Career Tools like MCA

- Must deliver **personalized, skills- and interest-based recommendations** that are relevant to local realities and specific companies or industries.
- Should include:
    - **Strengths and weaknesses analysis**.
    - **Learning paths aligned with industry standards**.
    - Guidance for those uncertain about career labels or paths.
    - Integration of **resume review** and multiple input sources (not just personality tests).
    - Features for **networking** and **peer learning**.

9. Trust and Credibility Are Critical

- Trust factors: credible sources, expert backing, success stories/testimonials, and demonstrable accuracy of recommendations.

- Distrust factors: unclear recommendation logic, misalignment with user interests, insecure handling of personal data.

Participants were asked to reflect on their sense of agency, the influence of family and media, and their trust in digital tools like MCA. They were also invited to critique the draft job list and share what kinds of work they felt were missing, misunderstood, or undervalued. This process ensured the tool aligns not just with labor market data, but with the values, dreams, and realities of today's youth.

Limitations and Future Considerations

Limitations

Future Updates

→ Plans for updating the job list and incorporating new data sources over time.

Contact and Version Control

- **Lead Implementer:** Wadhwani Foundation Philippines
- **Contact Us**: Edward.Landoy@WadhwaniFoundation.org
- **MCA Version History:**
  - **India - v1.0 (March 2025):** Pilot in India
  - **India - v1.1 (August 2025):** Scale-up in India
  - **Philippines - V1.0 (Q4 2026):** In development