

Term Project: Is AI taking our jobs or transforming them?

Lana Geissinger

Bellevue University

DSC540_T303 Data Preparation (2257-1)

Professor Catherine Williams

Milestone 4

Analyzing Occupation Skills Through O*NET Data

July 27, 2025

Connecting to an API/Pulling in the Data and Cleaning/Formatting

Milestone 4 focuses on connecting to the O*NET Webservices API to gather detailed occupation data, specifically focusing on skills and requirements for jobs identified in Milestone 3 "Growing/Declining occupations". The goal is to understand how job requirements and skills are evolving in response to AI.

```
import json
import pandas as pd
import sys
from html5lib.treeadapters.sax import prefix
from thefuzz import fuzz
import re
from typing import Dict, List
import time
import os
import requests
from dotenv import load_dotenv
from datetime import datetime
```

Connecting to O*NET Web Services API

Connection Setup

The ONET Web Services implementation is based on the official O*NET Center sample code: [O*NET Web Services Python Sample Code \(https://github.com/onetcenter/web-services-samples/blob/master/python-2/OnetWebService.py\)](https://github.com/onetcenter/web-services-samples/blob/master/python-2/OnetWebService.py).

- Add scripts directory to Python path for module access using
- Load environment variables for API authentication
- Test connection to O*NET Web Services

Configuration Notes

- **API Endpoint:** <https://services.onetcenter.org/ws/online/search>
- **Authentication:** Username/password from environment variables
- **Request Headers:**
 - User-Agent: Python/OnetWebService
 - Accept: application/json
 - Content-Type: application/json

Connection Test Process

- Load credentials from environment file
- Verify credentials exist
- Test connection with sample search
- Check response status (200 for success)
- Handle errors and provide feedback

```
# Add scripts directory to Python path
scripts_dir = os.path.abspath(os.path.join(os.path.dirname('OnetWebService.py'), '..', 'scripts'))
sys.path.append(scripts_dir)
```

```

from OnetWebService import OnetWebService

def test_onet_connection():

    print("\n🔑 Loading environment variables...")

    load_dotenv('../env_var.env')

    username = os.getenv('ONET_API_USERNAME')
    password = os.getenv('ONET_API_PASSWORD')

    # Verify credentials exist
    if not username or not password:
        print("❌ Missing O*NET API credentials in environment variables")
        print("Please set ONET_API_USERNAME and ONET_API_PASSWORD in env_var.env")
        return False

    print("✅ Found credentials:")
    print(f"Username: {'*' * len(username)}")
    print(f>Password: {'*' * len(password)}")

    # Test connection
    print("\n🌐 Testing connection to O*NET Web Services...")

    base_url = "https://services.onetcenter.org/ws/online/search" # Updated endpoint
    headers = {
        "User-Agent": "Python/OnetWebService",
        "Accept": "application/json",
        "Content-Type": "application/json"
    }

    # Test connection with simple search
    params = {
        "keyword": "software developer"
    }

    try:
        response = requests.get(
            base_url,
            auth=(username, password),
            headers=headers,
            params=params
        )

        if response.status_code == 200:
            print("✅ Successfully connected to O*NET Web Services")
            return True
        else:
            print(f"❌ Connection failed with status code: {response.status_code}")
            print(f"Response text: {response.text}")
            return False

    except Exception as e:
        print(f"❌ Error during connection test: {e}")
        return False

```

```
if __name__ == "__main__":
    test_onet_connection()
```

🔑 Loading environment variables...

✅ Found credentials:

Username: *****

Password: *****

🌐 Testing connection to O*NET Web Services...

✅ Successfully connected to O*NET Web Services

SOC Code Formatting for Growing_Declining.csv

This process starts by loading and cleaning occupation data from the Growing_Declining.csv file, which contains employment trends, and then matching those job titles with standardized classifications from the SOC_DB.csv file.

Once the titles are aligned between the two sources, SOC codes are added to the employment data to ensure consistency for analysis. After that, matrix codes are generated to align with BLS series or ONET formats. The final output is a new file called Growing_Declining_SOC.csv, which includes cleaned occupation titles, corresponding SOC codes, and the original employment matrix data.

```
# Load datasets
df_growing_declining = pd.read_csv('../output/Growing_Declining.csv')
df_soc = pd.read_csv('../output/SOC_DB.csv', encoding='Windows-1252')

# Clean and normalize titles
df_growing_declining['title_clean'] = df_growing_declining[
    '2023_national_employment_matrix_title'].str.lower().str.strip()
df_soc['title_clean'] = df_soc['occupation_title'].str.lower().str.strip()

# Merge datasets
df_merged = df_growing_declining.merge(
    df_soc[['detailed_occupation', 'occupation_title', 'title_clean']],
    on='title_clean',
    how='left'
)

# Cleanup and rename columns
df_merged = df_merged.rename(columns={
    'detailed_occupation': 'soc_code'
})
df_merged.drop(columns=['title_clean'], inplace=True)

# Save to CSV
df_merged.to_csv('../output/Growing_Declining_SOC.csv', index=False)
```

SOC Code Formatting for O*NET API

Convert BLS SOC codes to ONET format for API compatibility. The formatted codes enable proper ONET API queries for occupation data and employment trends analysis.

Format Changes

- Basic SOC code (11-9041) → O*NET format (11-9041.00)
- Add .00 suffix if no detail level
- Keep existing detail levels (.01, .02, etc.)

```

# Load the Growing_Declining dataset
df = pd.read_csv('../output/Growing_Declining_SOC.csv')

def format_soc_code_for_onet(soc_code):

    if pd.isna(soc_code):
        return None

    soc_code = str(soc_code).strip()

    if re.match(r'^\d{2}-\d{4}\.\d{2}$', soc_code):
        return soc_code

    if re.match(r'^\d{2}-\d{4}$', soc_code):
        return f"{soc_code}.00"

    if re.match(r'^\d{6}$', soc_code):
        return f"{soc_code[:2]}-{soc_code[2:6]}.00"

    return None

# Create a new column 'onet_soc_code' with O*NET formatted SOC codes
df['onet_soc_code'] = df['soc_code'].apply(format_soc_code_for_onet)

print("\nSOC code conversions:")
print(pd.concat([df['soc_code'], df['onet_soc_code']], axis=1).head(10))

# Count of successful conversions
total = len(df)
converted = df['onet_soc_code'].notna().sum()
print(f"\nConverted {converted} out of {total} SOC codes")

# Save updated dataset in output folder
output_file = '../output/Growing_Declining_ONET_SOC.csv'
df.to_csv(output_file, index=False)
print(f"\nSaved updated dataset to {output_file}")

# Display errors
failed = df[df['onet_soc_code'].isna()]
if not failed.empty:
    print("\nSOC codes that couldn't be converted:")
    print(failed[['soc_code', '2023_national_employment_matrix_title']])

```

SOC code conversions:

	soc_code	onet_soc_code
0	NaN	None
1	43-9022	43-9022.00
2	47-5043	47-5043.00
3	43-2011	43-2011.00
4	43-2021	43-2021.00
5	43-1011	43-1011.00
6	43-2011	43-2011.00
7	43-9021	43-9021.00
8	51-4071	51-4071.00
9	47-5044	47-5044.00

Converted 85 out of 88 SOC codes

Saved updated dataset to ../output/Growing_Declining_ONET_SOC.csv

SOC codes that couldn't be converted:

	soc_code	\
0	NaN	
44	NaN	
74	NaN	

	2023_national_employment_matrix_title
0	Total, All Occupations
44	Total, All Occupations
74	Substance Abuse, Behavioral Disorder, And Mental Health Counselors

O*NET Occupation Search

Using fuzzy matching search O*NET database for occupation titles and return dictionary mapping input titles to matched occupations.

Process

- Load occupation titles from dataset
- Query O*NET API with each title
- Score matches using fuzzy string comparison
- Filter and rank results
- Save matches to JSON for further analysis

Output

Creates a JSON file containing:

- Timestamp and metadata
- Matched occupations with scores
- Success/failure statistics

This file serves as an intermediate step in the data processing pipeline. It stores the results of fuzzy matching between occupation title and ONET database and contains metadata about the matching process. These matches will be used to retrieve occupational data.


```

# Add scripts directory to Python path
scripts_dir = os.path.abspath(os.path.join(os.path.dirname('OnetWebService.py'), '..', 'scripts'))
sys.path.append(scripts_dir)

def fuzzy_search_occupations(onet_ws: OnetWebService, occupation_titles: List[str],
                             min_score: int = 80, max_results: int = 3) -> Dict[str, List[dict]]:

    results = {}

    for title in occupation_titles:
        print(f"\nSearching for: {title}")

        try:
            search_response = onet_ws.call('online/search',
                                           keyword=title,
                                           end=10)

            if not search_response or 'occupation' not in search_response:
                print(f"✗ No results found for '{title}'")
                results[title] = []
                continue

            # Get all occupations from response
            occupations = search_response.get('occupation', [])
            if not isinstance(occupations, list):
                occupations = [occupations]

            # Score matches using fuzzy matching
            scored_matches = []
            for occ in occupations:
                if isinstance(occ, dict):
                    onet_title = occ.get('title', '')
                    score = fuzz.ratio(title.lower(), onet_title.lower())

                    if score >= min_score:
                        scored_matches.append({
                            'title': onet_title,
                            'code': occ.get('code', ''),
                            'match_score': score,
                            'href': occ.get('href', '')
                        })

            # Sort by score and take top matches
            scored_matches.sort(key=lambda x: x['match_score'], reverse=True)
            results[title] = scored_matches[:max_results]

            # Print matches
            if scored_matches:
                print("✅ Found matches:")
                for match in scored_matches[:max_results]:
                    print(f"  - {match['title']} (Score: {match['match_score']})")
            else:
                print("⚠️ No matches above minimum score threshold")

```

```
        time.sleep(1)

    except Exception as e:
        print(f"✖ Error processing '{title}': {str(e)}")
        results[title] = []
        continue

    return results
```

```

# Load occupation titles from Growing_Declining dataset
try:
    df = pd.read_csv('../output/Growing_Declining.csv')
    occupation_titles = df['2023_national_employment_matrix_title'].unique().tolist()

    # Initialize O*NET Web Service
    onet = OnetWebService()

    print(f"Processing {len(occupation_titles)} unique occupation titles...")

    # Process all occupations and store results
    all_results = fuzzy_search_occupations(onet, occupation_titles)

    # Prepare output data with metadata
    output_data = {
        "metadata": {
            "timestamp": datetime.now().strftime("%Y-%m-%d %H:%M:%S"),
            "total_occupations": len(occupation_titles),
            "min_match_score": 80,
            "max_results_per_occupation": 3
        },
        "matches": all_results
    }

    # Save all results to a single JSON file
    output_file = '../api_responses/onet_occupation_matches.json'
    with open(output_file, 'w') as f:
        json.dump(output_data, f, indent=2)

    print(f"\n✅ All matches saved to: {output_file}")

    # Print summary statistics
    matches_found = sum(len(matches) > 0 for matches in all_results.values())
    print(f"\nSummary:")
    print(f"- Total occupations processed: {len(occupation_titles)}")
    print(f"- Occupations with matches: {matches_found}")
    print(f"- Occupations without matches: {len(occupation_titles) - matches_found}")

except Exception as e:
    print(f"Error: {str(e)}")

```

Processing 61 unique occupation titles...

Searching for: Total, All Occupations

⚠ No matches above minimum score threshold

Searching for: Word Processors And Typists

✅ Found matches:

- Word Processors and Typists (Score: 100)

Searching for: Roof Bolters, Mining

✅ Found matches:

- Roof Bolters, Mining (Score: 100)

Searching for: Telephone Operators

✅ Found matches:

- Telephone Operators (Score: 100)

Searching for: Switchboard Operators, Including Answering Service

✅ Found matches:

- Switchboard Operators, Including Answering Service (Score: 100)

Searching for: Data Entry Keyers

✅ Found matches:

- Data Entry Keyers (Score: 100)

Searching for: Foundry Mold And Coremakers

✅ Found matches:

- Foundry Mold and Coremakers (Score: 100)

Searching for: Loading And Moving Machine Operators, Underground Mining

✅ Found matches:

- Loading and Moving Machine Operators, Underground Mining (Score: 100)

Searching for: Patternmakers, Metal And Plastic

✅ Found matches:

- Patternmakers, Metal and Plastic (Score: 100)
- Model Makers, Metal and Plastic (Score: 83)

Searching for: Telemarketers

✅ Found matches:

- Telemarketers (Score: 100)

Searching for: Engine And Other Machine Assemblers

✅ Found matches:

- Engine and Other Machine Assemblers (Score: 100)

Searching for: Manufactured Building And Mobile Home Installers

✅ Found matches:

- Manufactured Building and Mobile Home Installers (Score: 100)

Searching for: Drilling And Boring Machine Tool Setters, Operators, And Tenders, Metal And Plastic

✅ Found matches:

- Drilling and Boring Machine Tool Setters, Operators, and Tenders, Metal and Plastic (Score: 100)
- Milling and Planing Machine Setters, Operators, and Tenders, Metal and Plastic (Score: 91)

- Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic (Score: 90)

Searching for: Prepress Technicians And Workers

✓ Found matches:

- Prepress Technicians and Workers (Score: 100)

Searching for: Cutters And Trimmers, Hand

✓ Found matches:

- Cutters and Trimmers, Hand (Score: 100)

Searching for: Order Clerks

✓ Found matches:

- Order Clerks (Score: 100)

Searching for: Grinding And Polishing Workers, Hand

✓ Found matches:

- Grinding and Polishing Workers, Hand (Score: 100)

Searching for: Refractory Materials Repairers, Except Brickmasons

✓ Found matches:

- Refractory Materials Repairers, Except Brickmasons (Score: 100)

Searching for: Model Makers, Metal And Plastic

✓ Found matches:

- Model Makers, Metal and Plastic (Score: 100)

- Patternmakers, Metal and Plastic (Score: 83)

Searching for: Print Binding And Finishing Workers

✓ Found matches:

- Print Binding and Finishing Workers (Score: 100)

Searching for: Camera And Photographic Equipment Repairers

✓ Found matches:

- Camera and Photographic Equipment Repairers (Score: 100)

Searching for: Aircraft Structure, Surfaces, Rigging, And Systems Assemblers

✓ Found matches:

- Aircraft Structure, Surfaces, Rigging, and Systems Assemblers (Score: 100)

Searching for: Timing Device Assemblers And Adjusters

✓ Found matches:

- Timing Device Assemblers and Adjusters (Score: 100)

Searching for: Forging Machine Setters, Operators, And Tenders, Metal And Plastic

✓ Found matches:

- Forging Machine Setters, Operators, and Tenders, Metal and Plastic (Score: 100)

- Plating Machine Setters, Operators, and Tenders, Metal and Plastic (Score: 94)

- Milling and Planing Machine Setters, Operators, and Tenders, Metal and Plastic (Score: 88)

Searching for: Structural Metal Fabricators And Fitters

✓ Found matches:

- Structural Metal Fabricators and Fitters (Score: 100)

Searching for: File Clerks

- ✓ Found matches:
 - File Clerks (Score: 100)

Searching for: Office Machine Operators, Except Computer

- ✓ Found matches:
 - Office Machine Operators, Except Computer (Score: 100)

Searching for: Payroll And Timekeeping Clerks

- ✓ Found matches:
 - Payroll and Timekeeping Clerks (Score: 100)

Searching for: Tellers

- ✓ Found matches:
 - Tellers (Score: 100)

Searching for: Door-To-Door Sales Workers, News And Street Vendors, And Related Workers

- ✓ Found matches:
 - Door-to-Door Sales Workers, News and Street Vendors, and Related Workers (Score: 100)

Searching for: New Accounts Clerks

- ✓ Found matches:
 - New Accounts Clerks (Score: 100)

Searching for: Wind Turbine Service Technicians

- ✓ Found matches:
 - Wind Turbine Service Technicians (Score: 100)

Searching for: Solar Photovoltaic Installers

- ✓ Found matches:
 - Solar Photovoltaic Installers (Score: 100)

Searching for: Nurse Practitioners

- ✓ Found matches:
 - Nurse Practitioners (Score: 100)

Searching for: Data Scientists

- ✓ Found matches:
 - Data Scientists (Score: 100)
 - Materials Scientists (Score: 80)

Searching for: Information Security Analysts

- ✓ Found matches:
 - Information Security Analysts (Score: 100)

Searching for: Medical And Health Services Managers

- ✓ Found matches:
 - Medical and Health Services Managers (Score: 100)

Searching for: Physician Assistants

- ✓ Found matches:
 - Physician Assistants (Score: 100)

Searching for: Computer And Information Research Scientists

✓ Found matches:

- Computer and Information Research Scientists (Score: 100)

Searching for: Physical Therapist Assistants

✓ Found matches:

- Physical Therapist Assistants (Score: 100)
- Physical Therapist Aides (Score: 83)

Searching for: Operations Research Analysts

✓ Found matches:

- Operations Research Analysts (Score: 100)

Searching for: Occupational Therapy Assistants

✓ Found matches:

- Occupational Therapy Assistants (Score: 100)
- Occupational Therapists (Score: 85)
- Occupational Therapy Aides (Score: 84)

Searching for: Actuaries

✓ Found matches:

- Actuaries (Score: 100)

Searching for: Financial Examiners

✓ Found matches:

- Financial Examiners (Score: 100)
- Financial Managers (Score: 81)

Searching for: Home Health And Personal Care Aides

⚠ No matches above minimum score threshold

Searching for: Veterinary Assistants And Laboratory Animal Caretakers

✓ Found matches:

- Veterinary Assistants and Laboratory Animal Caretakers (Score: 100)

Searching for: Veterinary Technologists And Technicians

✓ Found matches:

- Veterinary Technologists and Technicians (Score: 100)

Searching for: Logisticians

✓ Found matches:

- Logisticians (Score: 100)

Searching for: Veterinarians

✓ Found matches:

- Veterinarians (Score: 100)

Searching for: Substance Abuse, Behavioral Disorder, And Mental Health Counselors

⚠ No matches above minimum score threshold

Searching for: Epidemiologists

✓ Found matches:

- Epidemiologists (Score: 100)

Searching for: Health Specialties Teachers, Postsecondary

✓ Found matches:

- Health Specialties Teachers, Postsecondary (Score: 100)

Searching for: Speech-Language Pathologists

- ✓ Found matches:
 - Speech-Language Pathologists (Score: 100)
 - Speech-Language Pathology Assistants (Score: 88)

Searching for: Medical Equipment Repairers

- ✓ Found matches:
 - Medical Equipment Repairers (Score: 100)
 - Medical Equipment Preparers (Score: 96)

Searching for: Software Developers

- ✓ Found matches:
 - Software Developers (Score: 100)

Searching for: Nursing Instructors And Teachers, Postsecondary

- ✓ Found matches:
 - Nursing Instructors and Teachers, Postsecondary (Score: 100)

Searching for: Massage Therapists

- ✓ Found matches:
 - Massage Therapists (Score: 100)

Searching for: Computer And Information Systems Managers

- ✓ Found matches:
 - Computer and Information Systems Managers (Score: 100)

Searching for: Industrial Machinery Mechanics

- ✓ Found matches:
 - Industrial Machinery Mechanics (Score: 100)

Searching for: Personal Financial Advisors

- ✓ Found matches:
 - Personal Financial Advisors (Score: 100)

Searching for: Psychiatric Technicians

- ✓ Found matches:
 - Psychiatric Technicians (Score: 100)

- ✓ All matches saved to: ../api_responses/onet_occupation_matches.json

Summary:

- Total occupations processed: 61
- Occupations with matches: 58
- Occupations without matches: 3

```
type(all_results)
```

```
Out[207]: dict
```


Retrieving Occupation Data from O*NET

This code retrieves detailed occupation information from the O*NET Web Services API.

Data Collection Process

1. Loads environment variables for API authentication
2. Initializes O*NET Web Service connection
3. Reads the Growing_Declining_ONET_SOC.csv file containing occupation codes
4. For each occupation with a valid O*NET SOC code:
 - Retrieves basic occupation details
 - Gets additional occupation description
 - Stores the following information:
 - Original occupation title
 - O*NET code
 - Job description
 - Job family classification

Progress Tracking to stay informed about the progress and monitor API call usage and optimize process efficiency

- Monitors processing progress to confirm thae program is still running and not frozen
- Shows completion status every 5 occupations
- Includes error handling for failed API calls

Data Storage

- Saves collected data to JSON file with metadata
- Includes timestamp and total occupations processed
- Stores results in './api_responses/occupation_details.json'

```

# Load environment variables
load_dotenv('../env_var.env')
username = os.getenv('ONET_API_USERNAME')
password = os.getenv('ONET_API_PASSWORD')

# Initialize O*NET Web Service
onet = OnetWebService()

# Create a list to store occupation data
occupation_data = []

# Load the Growing_Declining_ONET_SOC.csv
df = pd.read_csv('../output/Growing_Declining_ONET_SOC.csv')

# Progress tracking
total = len(df[df['onet_soc_code'].notna()])
processed = 0

# Process each occupation
for _, row in df.iterrows():
    if pd.notna(row['onet_soc_code']):
        onet_code = row['onet_soc_code']
        occupation_title = row['2023_national_employment_matrix_title']

        try:
            occupation_details = onet.call(f'online/occupations/{onet_code}')

            if occupation_details:
                details = onet.call(f'online/occupations/{onet_code}/details')

                occupation_data.append({
                    'original_title': occupation_title,
                    'onet_code': onet_code,
                    'description': occupation_details.get('description', ''),
                    'job_family': occupation_details.get('job_family', {}).get('name', '')
                })

                processed += 1
                if processed % 5 == 0:
                    print(f"Progress: {processed}/{total} occupations processed")

        except Exception as e:
            print(f"Error processing {onet_code}")

time.sleep(1)

```

Progress: 5/85 occupations processed
Progress: 10/85 occupations processed
Progress: 15/85 occupations processed
Progress: 20/85 occupations processed
Progress: 25/85 occupations processed
Progress: 30/85 occupations processed
Progress: 35/85 occupations processed
Progress: 40/85 occupations processed
Progress: 45/85 occupations processed
Progress: 50/85 occupations processed
API request error: 422 Client Error: Unprocessable Entity for url: https://services.onetcenter.org/ws/online/occupations/11-0081.00
API request error: 422 Client Error: Unprocessable Entity for url: https://services.onetcenter.org/ws/online/occupations/11-0011.00
Progress: 55/85 occupations processed
Progress: 60/85 occupations processed
Progress: 65/85 occupations processed
Progress: 70/85 occupations processed
Progress: 75/85 occupations processed
API request error: 422 Client Error: Unprocessable Entity for url: https://services.onetcenter.org/ws/online/occupations/11-0013.00
API request error: 422 Client Error: Unprocessable Entity for url: https://services.onetcenter.org/ws/online/occupations/11-0021.00
Progress: 80/85 occupations processed

```
# Write all occupation data in JSON file
output_file = '../api_responses/occupation_details.json'
json_data = {
    'occupations': occupation_data,
    'metadata': {
        'total_occupations': len(occupation_data),
        'timestamp': datetime.now().isoformat()
    }
}

# Save to JSON file
with open(output_file, 'w', encoding='utf-8') as f:
    json.dump(json_data, f, indent=2, ensure_ascii=False)

print(f"\nProcessing complete. Data saved to {output_file}")
print(f"Total occupations processed: {len(occupation_data)}")
```

Processing complete. Data saved to ../api_responses/occupation_details.json
Total occupations processed: 81

```
type(occupation_data)
```

Out[210]: list

```
print(occupation_data)
```

[{'original_title': 'Word Processors And Typists', 'onet_code': '43-9022.00', 'description': 'Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.', 'job_family': ''}, {'original_title': 'Roof Bolters, Mining', 'onet_code': '47-5043.00', 'description': 'Operate machinery to install roof support bolts in underground mine.', 'job_family': ''}, {'original_title': 'Telephone Operators', 'onet_code': '43-2011.00', 'description': 'Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.', 'job_family': ''}, {'original_title': 'Telephone Operators', 'onet_code': '43-2021.00', 'description': 'Provide information by accessing alphabetical, geographical, or other directories. Assist customers with special billing requests, such as charges to a third party and credits or refunds for incorrectly dialed numbers or bad connections. May handle emergency calls and assist children or people with physical disabilities to make telephone calls.', 'job_family': ''}, {'original_title': 'Switchboard Operators, Including Answering Service', 'onet_code': '43-1011.00', 'description': 'Directly supervise and coordinate the activities of clerical and administrative support workers.', 'job_family': ''}, {'original_title': 'Switchboard Operators, Including Answering Service', 'onet_code': '43-2011.00', 'description': 'Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.', 'job_family': ''}, {'original_title': 'Data Entry Keyers', 'onet_code': '43-9021.00', 'description': 'Operate data entry device, such as keyboard or photo composing perforator. Duties may include verifying data and preparing materials for printing.', 'job_family': ''}, {'original_title': 'Foundry Mold And Coremakers', 'onet_code': '51-4071.00', 'description': 'Make or form wax or sand cores or molds used in the production of metal castings in foundries.', 'job_family': ''}, {'original_title': 'Loading And Moving Machine Operators, Underground Mining', 'onet_code': '47-5044.00', 'description': 'Operate underground loading or moving machine to load or move coal, ore, or rock using shuttle or mine car or conveyors. Equipment may include power shovels, hoisting engines equipped with cable-drawn scraper or scoop, or machines equipped with gathering arms and conveyor.', 'job_family': ''}, {'original_title': 'Patternmakers, Metal And Plastic', 'onet_code': '51-4062.00', 'description': 'Lay out, machine, fit, and assemble castings and parts to metal or plastic foundry patterns, core boxes, or match plates.', 'job_family': ''}, {'original_title': 'Telemarketers', 'onet_code': '41-9031.00', 'description': 'Sell business goods or services, the selling of which requires a technical background equivalent to a baccalaureate degree in engineering.', 'job_family': ''}, {'original_title': 'Telemarketers', 'onet_code': '41-9041.00', 'description': 'Solicit donations or orders for goods or services over the telephone.', 'job_family': ''}, {'original_title': 'Engine And Other Machine Assemblers', 'onet_code': '51-2023.00', 'description': 'Assemble or modify electromechanical equipment or devices, such as servomechanisms, gyros, dynamometers, magnetic drums, tape drives, brakes, control linkage, actuators, and appliances.', 'job_family': ''}, {'original_title': 'Engine And Other Machine Assemblers', 'onet_code': '51-2031.00', 'description': 'Construct, assemble, or rebuild machines, such as engines, turbines, and similar equipment used in such industries as construction, extraction, textiles, and paper manufacturing.', 'job_family': ''}, {'original_title': 'Manufactured Building And Mobile Home Installers', 'onet_code': '49-9095.00', 'description': 'Move or install mobile homes or prefabricated buildings.', 'job_family': ''}, {'original_title': 'Drilling And Boring Machine Tool Setters, Operators, And Tenders, Metal And Plastic', 'onet_code': '51-4032.00', 'description': 'Set up, operate, or tend drilling machines to drill, bore, ream, mill, or countersink metal or plastic work pieces.', 'job_family': ''}, {'original_title':

'Prepress Technicians And Workers', 'onet_code': '51-5111.00', 'description': 'Format and proof text and images submitted by designers and clients into finished pages that can be printed. Includes digital and photo typesetting. May produce printing plates.', 'job_family': ''}, {'original_title': 'Cutters And Trimmers, Hand', 'onet_code': '51-9031.00', 'description': 'Use hand tools or hand-held power tools to cut and trim a variety of manufactured items, such as a carpet, fabric, stone, glass, or rubber.', 'job_family': ''}, {'original_title': 'Order Clerks', 'onet_code': '43-4141.00', 'description': 'Interview persons desiring to open accounts in financial institutions. Explain account services available to prospective customers and assist them in preparing applications.', 'job_family': ''}, {'original_title': 'Order Clerks', 'onet_code': '43-4151.00', 'description': 'Receive and process incoming orders for materials, merchandise, classified ads, or services such as repairs, installations, or rental of facilities. Generally receives orders via mail, phone, fax, or other electronic means. Duties include informing customers of receipt, prices, shipping dates, and delays; preparing contracts; and handling complaints.', 'job_family': ''}, {'original_title': 'Grinding And Polishing Workers, Hand', 'onet_code': '51-9022.00', 'description': 'Grind, sand, or polish, using hand tools or hand-held power tools, a variety of metal, wood, stone, clay, plastic, or glass objects. Includes chippers, buffers, and finishers.', 'job_family': ''}, {'original_title': 'Refractory Materials Repairers, Except Brickmasons', 'onet_code': '49-9045.00', 'description': 'Build or repair equipment such as furnaces, kilns, cupolas, boilers, converters, ladles, soaking pits, and ovens, using refractory materials.', 'job_family': ''}, {'original_title': 'Model Makers, Metal And Plastic', 'onet_code': '51-4061.00', 'description': 'Set up and operate machines, such as lathes, milling and engraving machines, and jig borers to make working models of metal or plastic objects. Includes template makers.', 'job_family': ''}, {'original_title': 'Print Binding And Finishing Workers', 'onet_code': '51-5113.00', 'description': 'Bind books and other publications or finish printed products by hand or machine. May set up binding and finishing machines.', 'job_family': ''}, {'original_title': 'Camera And Photographic Equipment Repairers', 'onet_code': '49-9061.00', 'description': 'Repair and adjust cameras and photographic equipment, including commercial video and motion picture camera equipment.', 'job_family': ''}, {'original_title': 'Aircraft Structure, Surfaces, Rigging, And Systems Assemblers', 'onet_code': '51-1011.00', 'description': 'Directly supervise and coordinate the activities of production and operating workers, such as inspectors, precision workers, machine setters and operators, assemblers, fabricators, and plant and system operators. Excludes team or work leaders.', 'job_family': ''}, {'original_title': 'Aircraft Structure, Surfaces, Rigging, And Systems Assemblers', 'onet_code': '51-2011.00', 'description': 'Assemble, fit, fasten, and install parts of airplanes, space vehicles, or missiles, such as tails, wings, fuselage, bulkheads, stabilizers, landing gear, rigging and control equipment, or heating and ventilating systems.', 'job_family': ''}, {'original_title': 'Timing Device Assemblers And Adjusters', 'onet_code': '51-2051.00', 'description': 'Laminate layers of fiberglass on molds to form boat decks and hulls, bodies for golf carts, automobiles, or other products.', 'job_family': ''}, {'original_title': 'Timing Device Assemblers And Adjusters', 'onet_code': '51-2061.00', 'description': 'Perform precision assembling or adjusting, within narrow tolerances, of timing devices such as digital clocks or timing devices with electrical or electronic components.', 'job_family': ''}, {'original_title': 'Forging Machine Setters, Operators, And Tenders, Metal And Plastic', 'onet_code': '51-4022.00', 'description': 'Set up, operate, or tend forging machines to taper, shape, or form metal or plastic parts.', 'job_family': ''}, {'original_title': 'Structural Metal Fabricators And Fitters', 'onet_code': '51-2031.00', 'description': 'Construct, assemble, or rebuild machines, such as engines, turbines, and similar equipment used in such industries as constr

uction, extraction, textiles, and paper manufacturing.', 'job_family': ''}, {'original_title': 'Structural Metal Fabricators And Fitters', 'onet_code': '51-2041.00', 'description': 'Fabricate, position, align, and fit parts of structural metal products.', 'job_family': ''}, {'original_title': 'File Clerks', 'onet_code': '43-4061.00', 'description': 'Determine eligibility of persons applying to receive assistance from government programs and agency resources, such as welfare, unemployment benefits, social security, and public housing.', 'job_family': ''}, {'original_title': 'File Clerks', 'onet_code': '43-4071.00', 'description': 'File correspondence, cards, invoices, receipts, and other records in alphabetical or numerical order or according to the filing system used. Locate and remove material from file when requested.', 'job_family': ''}, {'original_title': 'Office Machine Operators, Except Computer', 'onet_code': '43-9061.00', 'description': 'Perform duties too varied and diverse to be classified in any specific office clerical occupation, requiring knowledge of office systems and procedures. Clerical duties may be assigned in accordance with the office procedures of individual establishments and may include a combination of answering telephones, bookkeeping, typing or word processing, office machine operation, and filing.', 'job_family': ''}, {'original_title': 'Office Machine Operators, Except Computer', 'onet_code': '43-9071.00', 'description': 'Operate one or more of a variety of office machines, such as photocopying, photographic, and duplicating machines, or other office machines.', 'job_family': ''}, {'original_title': 'Payroll And Timekeeping Clerks', 'onet_code': '43-3041.00', 'description': 'In a gambling establishment, conduct financial transactions for patrons. Accept patron's credit application and verify credit references to provide check-cashing authorization or to establish house credit accounts. May reconcile daily summaries of transactions to balance books. May sell gambling chips, tokens, or tickets to patrons, or to other workers for resale to patrons. May convert gambling chips, tokens, or tickets to currency upon patron's request. May use a cash register or computer to record transaction.', 'job_family': ''}, {'original_title': 'Payroll And Timekeeping Clerks', 'onet_code': '43-3051.00', 'description': 'Compile and record employee time and payroll data. May compute employees' time worked, production, and commission. May compute and post wages and deductions, or prepare paychecks.', 'job_family': ''}, {'original_title': 'Tellers', 'onet_code': '43-3061.00', 'description': 'Compile information and records to draw up purchase orders for procurement of materials and services.', 'job_family': ''}, {'original_title': 'Tellers', 'onet_code': '43-3071.00', 'description': 'Receive and pay out money. Keep records of money and negotiable instruments involved in a financial institution's various transactions.', 'job_family': ''}, {'original_title': 'Door-To-Door Sales Workers, News And Street Vendors, And Related Workers', 'onet_code': '41-9091.00', 'description': 'Sell goods or services door-to-door or on the street.', 'job_family': ''}, {'original_title': 'New Accounts Clerks', 'onet_code': '43-4131.00', 'description': 'Interview loan applicants to elicit information; investigate applicants' backgrounds and verify references; prepare loan request papers; and forward findings, reports, and documents to appraisal department. Review loan papers to ensure completeness, and complete transactions between loan establishment, borrowers, and sellers upon approval of loan.', 'job_family': ''}, {'original_title': 'New Accounts Clerks', 'onet_code': '43-4141.00', 'description': 'Interview persons desiring to open accounts in financial institutions. Explain account services available to prospective customers and assist them in preparing applications.', 'job_family': ''}, {'original_title': 'Wind Turbine Service Technicians', 'onet_code': '49-9071.00', 'description': 'Perform work involving the skills of two or more maintenance or craft occupations to keep machines, mechanical equipment, or the structure of a building in repair. Duties may involve pipe fitting; HVAC maintenance; insulating; welding; machining; carpentry; repairing electrical or mechanical equipment; installing, aligning, and balancing new e

quipment; and repairing buildings, floors, or stairs.', 'job_family': ''}, {'original_title': 'Wind Turbine Service Technicians', 'onet_code': '49-9081.00', 'description': 'Inspect, diagnose, adjust, or repair wind turbines. Perform maintenance on wind turbine equipment including resolving electrical, mechanical, and hydraulic malfunctions.', 'job_family': ''}, {'original_title': 'Solar Photovoltaic Installers', 'onet_code': '47-2221.00', 'description': 'Raise, place, and unite iron or steel girders, columns, and other structural members to form completed structures or structural frameworks. May erect metal storage tanks and assemble prefabricated metal buildings.', 'job_family': ''}, {'original_title': 'Solar Photovoltaic Installers', 'onet_code': '47-2231.00', 'description': 'Assemble, install, or maintain solar photovoltaic (PV) systems on roofs or other structures in compliance with site assessment and schematics. May include measuring, cutting, assembling, and bolting structural framing and solar modules. May perform minor electrical work such as current checks.', 'job_family': ''}, {'original_title': 'Nurse Practitioners', 'onet_code': '29-1161.00', 'description': 'Diagnose and coordinate all aspects of the birthing process, either independently or as part of a healthcare team. May provide well-woman gynecological care. Must have specialized, graduate nursing education.', 'job_family': ''}, {'original_title': 'Nurse Practitioners', 'onet_code': '29-1171.00', 'description': 'Diagnose and treat acute, episodic, or chronic illness, independently or as part of a healthcare team. May focus on health promotion and disease prevention. May order, perform, or interpret diagnostic tests such as lab work and x rays. May prescribe medication. Must be registered nurses who have specialized graduate education.', 'job_family': ''}, {'original_title': 'Data Scientists', 'onet_code': '15-2041.00', 'description': 'Develop or apply mathematical or statistical theory and methods to collect, organize, interpret, and summarize numerical data to provide usable information. May specialize in fields such as biostatistics, agricultural statistics, business statistics, or economic statistics. Includes mathematical and survey statisticians.', 'job_family': ''}, {'original_title': 'Data Scientists', 'onet_code': '15-2051.00', 'description': 'Develop and implement a set of techniques or analytics applications to transform raw data into meaningful information using data-oriented programming languages and visualization software. Apply data mining, data modeling, natural language processing, and machine learning to extract and analyze information from large structured and unstructured datasets. Visualize, interpret, and report data findings. May create dynamic data reports.', 'job_family': ''}, {'original_title': 'Information Security Analysts', 'onet_code': '15-1212.00', 'description': 'Plan, implement, upgrade, or monitor security measures for the protection of computer networks and information. Assess system vulnerabilities for security risks and propose and implement risk mitigation strategies. May ensure appropriate security controls are in place that will safeguard digital files and vital electronic infrastructure. May respond to computer security breaches and viruses.', 'job_family': ''}, {'original_title': 'Physician Assistants', 'onet_code': '29-1051.00', 'description': 'Dispense drugs prescribed by physicians and other health practitioners and provide information to patients about medications and their use. May advise physicians and other health practitioners on the selection, dosage, interactions, and side effects of medications.', 'job_family': ''}, {'original_title': 'Physician Assistants', 'onet_code': '29-1071.00', 'description': 'Provide healthcare services typically performed by a physician, under the supervision of a physician. Conduct complete physicals, provide treatment, and counsel patients. May, in some cases, prescribe medication. Must graduate from an accredited educational program for physician assistants.', 'job_family': ''}, {'original_title': 'Computer And Information Research Scientists', 'onet_code': '15-1212.00', 'description': 'Plan, implement, upgrade, or monitor security measures for the protection of computer networks and information. Assess system vulnerabilities for security risks and propose and imp

lement risk mitigation strategies. May ensure appropriate security controls are in place that will safeguard digital files and vital electronic infrastructure. May respond to computer security breaches and viruses.', 'job_family': '', {'original_title': 'Computer And Information Research Scientists', 'onet_code': '15-1221.00', 'description': 'Conduct research into fundamental computer and information science as theorists, designers, or inventors. Develop solutions to problems in the field of computer hardware and software.', 'job_family': ''}, {'original_title': 'Physical Therapist Assistants', 'onet_code': '31-2021.00', 'description': 'Assist physical therapists in providing physical therapy treatments and procedures. May, in accordance with state laws, assist in the development of treatment plans, carry out routine functions, document the progress of treatment, and modify specific treatments in accordance with patient status and within the scope of treatment plans established by a physical therapist. Generally requires formal training.', 'job_family': ''}, {'original_title': 'Operations Research Analysts', 'onet_code': '15-2021.00', 'description': 'Conduct research in fundamental mathematics or in application of mathematical techniques to science, management, and other fields. Solve problems in various fields using mathematical methods.', 'job_family': ''}, {'original_title': 'Operations Research Analysts', 'onet_code': '15-2031.00', 'description': 'Formulate and apply mathematical modeling and other optimizing methods to develop and interpret information that assists management with decisionmaking, policy formulation, or other managerial functions. May collect and analyze data and develop decision support software, services, or products. May develop and supply optimal time, cost, or logistics networks for program evaluation, review, or implementation.', 'job_family': ''}, {'original_title': 'Occupational Therapy Assistants', 'onet_code': '31-2011.00', 'description': 'Assist occupational therapists in providing occupational therapy treatments and procedures. May, in accordance with state laws, assist in development of treatment plans, carry out routine functions, direct activity programs, and document the progress of treatments. Generally requires formal training.', 'job_family': ''}, {'original_title': 'Actuaries', 'onet_code': '15-1299.00', 'description': 'All computer occupations not listed separately.', 'job_family': ''}, {'original_title': 'Actuaries', 'onet_code': '15-2011.00', 'description': 'Analyze statistical data, such as mortality, accident, sickness, disability, and retirement rates and construct probability tables to forecast risk and liability for payment of future benefits. May ascertain insurance rates required and cash reserves necessary to ensure payment of future benefits.', 'job_family': ''}, {'original_title': 'Financial Examiners', 'onet_code': '13-2054.00', 'description': 'Analyze and measure exposure to credit and market risk threatening the assets, earning capacity, or economic state of an organization. May make recommendations to limit risk.', 'job_family': ''}, {'original_title': 'Financial Examiners', 'onet_code': '13-2061.00', 'description': 'Enforce or ensure compliance with laws and regulations governing financial and securities institutions and financial and real estate transactions. May examine, verify, or authenticate records.', 'job_family': ''}, {'original_title': 'Home Health And Personal Care Aides', 'onet_code': '31-9099.00', 'description': 'All healthcare support workers not listed separately.', 'job_family': ''}, {'original_title': 'Veterinary Assistants And Laboratory Animal Care takers', 'onet_code': '31-9096.00', 'description': 'Feed, water, and examine pets and other nonfarm animals for signs of illness, disease, or injury in laboratories and animal hospitals and clinics. Clean and disinfect cages and work areas, and sterilize laboratory and surgical equipment. May provide routine postoperative care, administer medication orally or topically, or prepare samples for laboratory examination under the supervision of veterinary or laboratory animal technologists or technicians, veterinarians, or scientists.', 'job_family': ''}, {'original_title': 'Veterinary Technologists And Technicians', 'onet_code': '29-2056.00', 'description': 'Perform medical tests in a la

laboratory environment for use in the treatment and diagnosis of diseases in animals. Prepare vaccines and serums for prevention of diseases. Prepare tissue samples, take blood samples, and execute laboratory tests, such as urinalysis and blood counts. Clean and sterilize instruments and materials and maintain equipment and machines. May assist a veterinarian during surgery.', 'job_family': ''}, {'original_title': 'Logisticians', 'onet_code': '13-1081.00', 'description': 'Analyze and coordinate the ongoing logistical functions of a firm or organization. Responsible for the entire life cycle of a product, including acquisition, distribution, internal allocation, delivery, and final disposal of resources.', 'job_family': ''}, {'original_title': 'Veterinarians', 'onet_code': '29-1129.00', 'description': 'All therapists not listed separately.', 'job_family': ''}, {'original_title': 'Veterinarians', 'onet_code': '29-1131.00', 'description': 'Diagnose, treat, or research diseases and injuries of animals. Includes veterinarians who conduct research and development, inspect livestock, or care for pets and companion animals.', 'job_family': ''}, {'original_title': 'Epidemiologists', 'onet_code': '19-1041.00', 'description': 'Investigate and describe the determinants and distribution of disease, disability, or health outcomes. May develop the means for prevention and control.', 'job_family': ''}, {'original_title': 'Health Specialties Teachers, Postsecondary', 'onet_code': '25-1071.00', 'description': 'Teach courses in health specialties, in fields such as dentistry, laboratory technology, medicine, pharmacy, public health, therapy, and veterinary medicine.', 'job_family': ''}, {'original_title': 'Speech-Language Pathologists', 'onet_code': '29-1127.00', 'description': 'Assess and treat persons with speech, language, voice, and fluency disorders. May select alternative communication systems and teach their use. May perform research related to speech and language problems.', 'job_family': ''}, {'original_title': 'Medical Equipment Repairers', 'onet_code': '49-9062.00', 'description': 'Test, adjust, or repair biomedical or electromedical equipment.', 'job_family': ''}, {'original_title': 'Software Developers', 'onet_code': '15-1252.00', 'description': 'Research, design, and develop computer and network software or specialized utility programs. Analyze user needs and develop software solutions, applying principles and techniques of computer science, engineering, and mathematical analysis. Update software or enhance existing software capabilities. May work with computer hardware engineers to integrate hardware and software systems, and develop specifications and performance requirements. May maintain databases within an application area, working individually or coordinating database development as part of a team.', 'job_family': ''}, {'original_title': 'Nursing Instructors And Teachers, Postsecondary', 'onet_code': '25-1072.00', 'description': 'Demonstrate and teach patient care in classroom and clinical units to nursing students. Includes both teachers primarily engaged in teaching and those who do a combination of teaching and research.', 'job_family': ''}, {'original_title': 'Massage Therapists', 'onet_code': '31-2022.00', 'description': 'Under close supervision of a physical therapist or physical therapy assistant, perform only delegated, selected, or routine tasks in specific situations. These duties include preparing the patient and the treatment area.', 'job_family': ''}, {'original_title': 'Massage Therapists', 'onet_code': '31-9011.00', 'description': 'Perform therapeutic massages of soft tissues and joints. May assist in the assessment of range of motion and muscle strength, or propose client therapy plans.', 'job_family': ''}, {'original_title': 'Industrial Machinery Mechanics', 'onet_code': '49-9041.00', 'description': 'Repair, install, adjust, or maintain industrial production and processing machinery or refinery and pipeline distribution systems. May also install, dismantle, or move machinery and heavy equipment according to plans.', 'job_family': ''}, {'original_title': 'Personal Financial Advisors', 'onet_code': '13-2052.00', 'description': 'Advise clients on financial plans using knowledge of tax and investment strategies, securities, insurance, pension plans, and real estate. Duties include assessing

clients' assets, liabilities, cash flow, insurance coverage, tax status, and financial objectives. May also buy and sell financial assets for clients.", 'job_family': ''}, {'original_title': 'Psychiatric Technicians', 'onet_code': '29-2053.00', 'description': "Care for individuals with mental or emotional conditions or disabilities, following the instructions of physicians or other health practitioners. Monitor patients' physical and emotional well-being and report to medical staff. May participate in rehabilitation and treatment programs, help with personal hygiene, and administer oral or injectable medications.", 'job_family': ''}]

```
pd.json_normalize(occupation_data).head()
```

Out[212]:

	original_title	onet_code	description	job_family
0	Word Processors And Typists	43- 9022.00	Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.	
1	Roof Bolters, Mining	47- 5043.00	Operate machinery to install roof support bolts in underground mine.	
2	Telephone Operators	43-2011.00	Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.	
3	Telephone Operators	43- 2021.00	Provide information by accessing alphabetical, geographical, or other directories. Assist customers with special billing requests, such as charges to a third party and credits or refunds for incorrectly dialed numbers or bad connections. May handle emergency calls and assist children or people with physical disabilities to make telephone calls.	
4	Switchboard Operators, Including Answering Service	43-1011.00	Directly supervise and coordinate the activities of clerical and administrative support workers.	

```

# Load the growing_declining_ONET_SOC file
df_growing_declining = pd.read_csv('../output/Growing_Declining_ONET_SOC.csv')

# Get the occupation data
with open('../api_responses/occupation_details.json', 'r', encoding='utf-8') as f:
    occupation_data = json.load(f)

# Convert occupation_data to DataFrame
df_occupations = pd.json_normalize(occupation_data['occupations'])

# Rename the column in df_growing_declining
df_growing_declining = df_growing_declining.rename(columns={'onet_soc_code': 'onet_code'})

# Add columns from growing_declining to occupation_data
for occ in occupation_data['occupations']:
    matches = df_growing_declining[df_growing_declining['onet_code'] == occ['onet_code']]
    if not matches.empty:
        record = matches.iloc[0]
        occ.update({
            '2023_national_employment_matrix_title': record['2023_national_employment_matrix_title'],
            'growth_status': record['growth_status'],
            'soc_code': record['soc_code']
        })
    else:
        occ.update({
            '2023_national_employment_matrix_title': None,
            'growth_status': None,
            'soc_code': None
        })

# Save the updated occupation_details.json
with open('../api_responses/occupation_details_upd.json', 'w', encoding='utf-8') as f:
    json.dump(occupation_data, f, indent=2, ensure_ascii=False)

```

```
pd.json_normalize(occupation_data['occupations']).head()
```

Out[214]:

	original_title	onet_code	description	job_family	2023_national_employment_matrix_title	grow
0	Word Processors And Typists	43-9022.00	Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.		Word Processors And Typists	
1	Roof Bolters, Mining	47-5043.00	Operate machinery to install roof support bolts in underground mine.		Roof Bolters, Mining	
2	Telephone Operators	43-2011.00	Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.		Telephone Operators	

original_title	onet_code	description	job_family	2023_national_employment_matrix_title	grow
3	Telephone Operators	43-2021.00	Provide information by accessing alphabetical, geographical, or other directories. Assist customers with special billing requests, such as charges to a third party and credits or refunds for incorrectly dialed numbers or bad connections. May handle emergency calls and assist children or people with physical disabilities to make telephone calls.	Telephone Operators	
4	Switchboard Operators, Including Answering Service	43-1011.00	Directly supervise and coordinate the activities of clerical and administrative support workers.	Switchboard Operators, Including Answering Service	

```
# Read the JSON file
with open('../api_responses/occupation_details_upd.json', 'r', encoding='utf-8') as f:
    occupation_data = json.load(f)

# Create DataFrame from the occupations List in the JSON
df_occupations = pd.json_normalize(occupation_data['occupations'])

# Display the first few rows and basic information about the DataFrame
print("DataFrame Preview:")
print(df_occupations.head())
print("\nDataFrame Info:")
print(df_occupations.info())
```


DataFrame Preview:

	original_title	onet_code	\
0	Word Processors And Typists	43-9022.00	
1	Roof Bolters, Mining	47-5043.00	
2	Telephone Operators	43-2011.00	
3	Telephone Operators	43-2021.00	
4	Switchboard Operators, Including Answering Service	43-1011.00	

description \

0
Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.

1
Operate machinery to install roof support bolts in underground mine.

2
Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.

3 Provide information by accessing alphabetical, geographical, or other directories. Assist customers with special billing requests, such as charges to a third party and credits or refunds for incorrectly dialed numbers or bad connections. May handle emergency calls and assist children or people with physical disabilities to make telephone calls.

4
Directly supervise and coordinate the activities of clerical and administrative support workers.

job_family	2023_national_employment_matrix_title	\
0	Word Processors And Typists	
1	Roof Bolters, Mining	
2	Telephone Operators	
3	Telephone Operators	
4	Switchboard Operators, Including Answering Service	

	growth_status	soc_code
0	Declining	43-9022
1	Declining	47-5043
2	Declining	43-2011
3	Declining	43-2021
4	Declining	43-1011

DataFrame Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 81 entries, 0 to 80

Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype
0	original_title	81 non-null	object
1	onet_code	81 non-null	object
2	description	81 non-null	object
3	job_family	81 non-null	object
4	2023_national_employment_matrix_title	81 non-null	object
5	growth_status	81 non-null	object
6	soc_code	81 non-null	object

dtypes: object(7)

memory usage: 4.6+ KB
None

Transformation: Step 1 - Replace Headers

```
column_mapping = {  
    'onet_code': 'onet_soc_code',  
    'original_title': 'occupation_title',  
    'description': 'occupation_description',  
    '2023_national_employment_matrix_title': 'employment_matrix_title',  
    'growth_status': 'employment_trend'  
}  
df_occupations = df_occupations.rename(columns=column_mapping)
```

```
df_occupations.head()
```

Out[217]:

	occupation_title	onet_soc_code	occupation_description	job_family	employment_matrix_title	ε
0	Word Processors And Typists	43-9022.00	Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.		Word Processors And Typists	
1	Roof Bolters, Mining	47-5043.00	Operate machinery to install roof support bolts in underground mine.		Roof Bolters, Mining	
2	Telephone Operators	43-2011.00	Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.		Telephone Operators	
3	Telephone Operators	43-2021.00	Provide information by accessing alphabetical, geographical, or other directories. Assist customers with special billing requests, such as charges to a third party and credits or refunds for incorrectly dialed numbers or bad connections. May handle emergency calls and assist children or people with physical disabilities to make telephone calls.		Telephone Operators	
4	Switchboard Operators, Including Answering Service	43-1011.00	Directly supervise and coordinate the activities of clerical and administrative support workers.		Switchboard Operators, Including Answering Service	

Transformation: Step 2 - Find and Remove Duplicates

```
print("Number of duplicate titles:",  
      df_occupations['occupation_title'].duplicated().sum())  
print(df_occupations[df_occupations['occupation_title'].duplicated(keep=False)]  
      .sort_values('occupation_title')[['occupation_title', 'onet_soc_code']])
```

Number of duplicate titles: 24

	occupation_title \
60	Actuaries
61	Actuaries
25	Aircraft Structure, Surfaces, Rigging, And Systems Assemblers
26	Aircraft Structure, Surfaces, Rigging, And Systems Assemblers
55	Computer And Information Research Scientists
54	Computer And Information Research Scientists
50	Data Scientists
49	Data Scientists
12	Engine And Other Machine Assemblers
13	Engine And Other Machine Assemblers
33	File Clerks
32	File Clerks
63	Financial Examiners
62	Financial Examiners
76	Massage Therapists
77	Massage Therapists
42	New Accounts Clerks
41	New Accounts Clerks
48	Nurse Practitioners
47	Nurse Practitioners
34	Office Machine Operators, Except Computer
35	Office Machine Operators, Except Computer
57	Operations Research Analysts
58	Operations Research Analysts
18	Order Clerks
19	Order Clerks
37	Payroll And Timekeeping Clerks
36	Payroll And Timekeeping Clerks
52	Physician Assistants
53	Physician Assistants
45	Solar Photovoltaic Installers
46	Solar Photovoltaic Installers
31	Structural Metal Fabricators And Fitters
30	Structural Metal Fabricators And Fitters
5	Switchboard Operators, Including Answering Service
4	Switchboard Operators, Including Answering Service
11	Telemarketers
10	Telemarketers
2	Telephone Operators
3	Telephone Operators
38	Tellers
39	Tellers
28	Timing Device Assemblers And Adjusters
27	Timing Device Assemblers And Adjusters
68	Veterinarians
69	Veterinarians
43	Wind Turbine Service Technicians
44	Wind Turbine Service Technicians

	onet_soc_code
60	15-1299.00
61	15-2011.00
25	51-1011.00
26	51-2011.00
55	15-1221.00

54	15-1212.00
50	15-2051.00
49	15-2041.00
12	51-2023.00
13	51-2031.00
33	43-4071.00
32	43-4061.00
63	13-2061.00
62	13-2054.00
76	31-2022.00
77	31-9011.00
42	43-4141.00
41	43-4131.00
48	29-1171.00
47	29-1161.00
34	43-9061.00
35	43-9071.00
57	15-2021.00
58	15-2031.00
18	43-4141.00
19	43-4151.00
37	43-3051.00
36	43-3041.00
52	29-1051.00
53	29-1071.00
45	47-2221.00
46	47-2231.00
31	51-2041.00
30	51-2031.00
5	43-2011.00
4	43-1011.00
11	41-9041.00
10	41-9031.00
2	43-2011.00
3	43-2021.00
38	43-3061.00
39	43-3071.00
28	51-2061.00
27	51-2051.00
68	29-1129.00
69	29-1131.00
43	49-9071.00
44	49-9081.00

```
df_occupations_unique = df_occupations.drop_duplicates(  
    subset=['occupation_title'],  
    keep='first'  
).reset_index(drop=True)
```

```
df_occupations_unique.head()
```

Out[220]:

	occupation_title	onet_soc_code	occupation_description	job_family	employment_matrix_title	ε
0	Word Processors And Typists	43-9022.00	Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.		Word Processors And Typists	
1	Roof Bolters, Mining	47-5043.00	Operate machinery to install roof support bolts in underground mine.		Roof Bolters, Mining	
2	Telephone Operators	43-2011.00	Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.		Telephone Operators	
3	Switchboard Operators, Including Answering Service	43-1011.00	Directly supervise and coordinate the activities of clerical and administrative support workers.		Switchboard Operators, Including Answering Service	
4	Data Entry Keyers	43-9021.00	Operate data entry device, such as keyboard or photo composing perforator. Duties may include verifying data and preparing materials for printing.		Data Entry Keyers	

Transformation: Step 3 - Check for missing values, empty strings and whitespace in column

```
# Function to analyze empty and whitespace values in a column
def analyze_column(df, column):
    if df[column].dtype == object:
        empty_strings = (df[column] == '').sum()
        whitespace_only = df[column].str.isspace().sum() if df[column].dtype == object else 0
        leading_trailing = (df[column].str.strip() != df[column]).sum() if df[column].dtype == object else 0
        null_values = df[column].isna().sum()

        print(f"\nColumn: {column}")
        print(f"Empty strings (''): {empty_strings}")
        print(f"Whitespace-only strings: {whitespace_only}")
        print(f"Leading/trailing whitespace: {leading_trailing}")
        print(f"NULL values: {null_values}")

# Analyze each column in the dataframe
print("Analysis of empty strings and whitespace by column:")
for column in df_occupations_unique.columns:
    analyze_column(df_occupations_unique, column)
```


Analysis of empty strings and whitespace by column:

Column: occupation_title
Empty strings (''): 0
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Column: onet_soc_code
Empty strings (''): 0
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Column: occupation_description
Empty strings (''): 0
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Column: job_family
Empty strings (''): 57
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Column: employment_matrix_title
Empty strings (''): 0
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Column: employment_trend
Empty strings (''): 0
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Column: soc_code
Empty strings (''): 0
Whitespace-only strings: 0
Leading/trailing whitespace: 0
NULL values: 0

Transformation: Step 4 - Remove column 'job_family' with missing values

```
df_occupations_unique = df_occupations_unique.drop(columns=['job_family'])
```

```
df_occupations_unique.head()
```

Out[223]:

	occupation_title	onet_soc_code	occupation_description	employment_matrix_title	employment_
0	Word Processors And Typists	43-9022.00	Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.	Word Processors And Typists	Der
1	Roof Bolters, Mining	47-5043.00	Operate machinery to install roof support bolts in underground mine.	Roof Bolters, Mining	Der
2	Telephone Operators	43-2011.00	Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.	Telephone Operators	Der
3	Switchboard Operators, Including Answering Service	43-1011.00	Directly supervise and coordinate the activities of clerical and administrative support workers.	Switchboard Operators, Including Answering Service	Der
4	Data Entry Keyers	43-9021.00	Operate data entry device, such as keyboard or photo composing perforator. Duties may include verifying data and preparing materials for printing.	Data Entry Keyers	Der

Transformation: Step 5 - Reorder columns

[illegible]

```
df_occupations_unique.head()
```

Out[225]:

	occupation_title	employment_trend	onet_soc_code	soc_code	occupation_description
0	Word Processors And Typists	Declining	43-9022.00	43-9022	Use word processor, computer, or typewriter to type letters, reports, forms, or other material from rough draft, corrected copy, or voice recording. May perform other clerical duties as assigned.
1	Roof Bolters, Mining	Declining	47-5043.00	47-5043	Operate machinery to install roof support bolts in underground mine.
2	Telephone Operators	Declining	43-2011.00	43-2011	Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.
3	Switchboard Operators, Including Answering Service	Declining	43-1011.00	43-1011	Directly supervise and coordinate the activities of clerical and administrative support workers.
4	Data Entry Keyers	Declining	43-9021.00	43-9021	Operate data entry device, such as keyboard or photo composing perforator. Duties may include verifying data and preparing materials for printing.

```
# Save the cleaned file to output folder for loading into SQL DB in Milestone 5
```

```
# Output file path
```

```
output_dir = os.path.join '..', 'output')
```

```
output_file = os.path.join(output_dir, 'Occupation_Descriptions.csv')
```

```
# Save as CSV
```

```
df_occupations_unique.to_csv(output_file, index=False)
```

```
# Verify the file was created
```

```
if os.path.exists(output_file):
```

```
    print(f"File successfully saved to: {output_file}")
```

```
else:
```

```
    print("Error: File was not created")
```

File successfully saved to: ..\output\Occupation_Descriptions.csv

Ethical Implications Of Connecting to API and cleaning/formatting ONET occupation data

While connecting to API and cleaning/formatting ONET occupation data, I performed the following steps:

Initial Setup and Connection Testing:

- Configured ONET Web Services connection using official sample code
- Implemented secure credential management using environment variables
- Tested API connection
- Validated response status and error handling
- Added rate limiting to respect API constraints

Data Processing Setup and SOC Code formatting:

- Loaded the Growing_Declining_SOC.scv file with occupational data
- Implemented SOC code formatting function for ONET compatibility
- Validated and saved formatted codes to Growing_Declining_ONET_SOC.csv

ONET occupation data cleaning and formatting steps to occupation_details.json file:

- Implemented fuzzy matching search for occupation titles
 - Queried ONET API for each occupation with formatted SOC codes and scored and ranked matches
 - Collected occupation data and saved results in occupation_details.json
 - Normalized json data using panda
 - Add columns from modified growing_declining.csv and saved results in occupation_details_upd.json
- In occupation_details.json:*
- Replaced headers
 - Found and removed duplicates
 - Checked for missing values, empty strings and whitespaces in all columns
 - Removed colum 'job_family' with all values missing
 - Reordered columns
 - Saved the cleaned file to the output folder for loading into SQL DB in Milestone 5

Ethical Implications:

All data came from public sources like BLS and O*NET, keeping it legal and ethical for research use.

Growth categories and metrics were clearly labeled to avoid confusion around workforce trends.

Some bias may come from assumptions during fuzzy matching and cleaning steps.

To help with that, I used multiple sources and documented the process.

The project doesn't make firm predictions about AI's impact but stays transparent and notes its limits.