# Objective

The objective of this take home project is to assess your data science, data engineering, NLP and coding skills.

As part of the project, you will pre-process data, apply NLP techniques, and gather insights with a clean, readable code.

You will be identifying and analyzing relevant entities from scraped job postings data. The relevant entities for the purpose of this take-home project are limited to:

1. **Skills**: such as “writing, critical thinking, persuasion” etc.
2. **Technology Skills**: such as “python, R, SAP” etc.

# Data

There are 6 csv files, containing job postings data from November 2019, October 2020, and October 2021 for Massachusetts (MA) and Missouri (MO). Data includes following fields:

* Job ID
* Job Title
* SOC Code (you can ignore it for the purpose of this exercise)
* Job Description
* Company Name
* Skills ( you can use this as well)
* Qualification
* City, State,Zipcode
* Job Opening / Close Dates
* Status and URL

There are 2 excel files, containing taxonomy data for soft and hard skills.

* O’Net Skills Taxonomy:
  + O\*NET-SOC Code
  + Title (SOC Title)
  + Element ID
  + Element Name (Skills)
  + Scale ID
  + Scale Name
  + Data Value
  + N
  + Standard Error
  + Lower CI Bound
  + Upper CI Bound
  + Recommend Suppress
  + Not Relevant
  + Date
  + Domain Source
* O’Net Technology Skills Taxonomy:
  + O\*NET-SOC Code
  + Title (SOC)
  + Example (Technology Skills)
  + Commodity Code
  + Commodity Title
  + Hot Technology

Primarily features that you want to use to accomplish your tasks are highlighted in yellow. As time permits, you may use other features as an enhancement.

# Expected Output:

1. **Exploratory Data Analysis - Raw Data**

*(this assesses your critical thinking and data science skills)*

We would like you to explore the job postings data before you begin the analysis. You will need to apply some pre-processing for this analysis.

* Have you noticed anything unusual about the data? What are your insights?
* What does the data look like? What type of cleaning / preprocessing would you do on the data if you have more time? Identify and explain the tools you will use in this analysis.
* Other Insights?

1. **Entity Extraction and Categorization**

*(this assesses your NLP and data engineering skills)*

You can use any approach to solve this problem. The more weights/points will be awarded to solutions that do broad matching (e.g dictionary says “machine learning” as skill, but you are able to find “machine-learning”, “ml”, “machine lerning”, “ML”) .

| Job ID | Entity Type | Entity Values |
| --- | --- | --- |
| XYZ1 | Skills | [Programming, Troubleshooting, Time Management, Social Perceptiveness...] |
| XYZ1 | Technology Skills | [Microsoft Access, SAP, Airtable, Marketo Marketing Automation...] |
| ... | ... | ... |

*Note: This is just a take-home project, we are not expecting you to have close to 100% coverage for entity extraction -- do best you can given time constraints.*

1. **Summary Statistics**

*(this assesses your data engineering skills)*

| Entity | Entity Type | Count | Location | Date |
| --- | --- | --- | --- | --- |
| Programming | Skills | 30 | MA | Oct 2020 |
| Airtable | Technology Skills | 423 | MO | Oct 2021 |
| Marketing Automation | Technology Skills | 525 | MO | Oct 2020 |

*Note: numbers above are random for illustration purposes*

1. **Insights - Post-Processed Data**

*(this assesses your critical thinking and data science skills)*

Can you do some additional exploratory analysis with this new data? Some starter questions:

1. What changed between November 2019, Oct 2020, and Oct 2021 for the job market in those markets?
   1. Which skills that were popular in 2019 or 2020 are no longer popular in 2021?
   2. Which skills gained or lost popularity?
2. Are there any apparent employment type differences between MA and MO markets?
3. Top year over year trends (trendy skills -bot down and up) by market
4. Other

*Note: We would like to see figures with good insight.*

1. **Create Similarity Score**

*(this assesses your NLP skills)*

Use job descriptions as documents, calculate pairwise document similarity scores for all jobs from October 2021 in MA only (this could be a resource intensive task). You are expected to use an NLP approach of your choice

Your approach does not have to be an optimum solution, you can provide an answer to what an optimum solution looks like.

| Job ID A | Job ID B | Similarity Score |
| --- | --- | --- |
| XYZ1 | XYZ2 | 0.874 |
| XYZ1 | XYZ3 | 0.756 |
| ... | ... | ... |

1. **Code + Output Data**

*(this assesses your coding skills)*

Please provide:

1. Readable code of your choice (python is preferable) for the above tasks
   1. A jupyter notebook would be the best solution since for task #3 we can see the plots (please add comments for insights as well)
2. Scalable for bigger datasets

*Note: don’t worry about parallel jobs or optimizing the code. You can use a for loop if you think it fits this task.*

1. Output Data in CSV format (for tables)

*Note: We would like to see some quality in the code that could be a part of a scalable data pipeline.*

# Sample Job Posting

| Field | Value |
| --- | --- |
| Job Id | e87d999a042270d971596679415b1376 |
| Job Title | Manager, Broadcast Operations |
| SOC Code | 11-1021.00 |
| Job Description | <br/> <br/>The TOC (Technical Operations Center) is a television programming environment where the operations team is responsible for monitoring, responding, escalating, documenting and maintaining the quality of all services that are re-distributed to its customers. Communication in written and verbal form with station personnel, internal departments and co-workers is essential to the performance of the team. All TOC personnel are expected to be able to multi-task, pay attention to fine details, demonstrate good teamwork and follow instructions given to them in an orderly and timely manner. Each member must have the ability to manage their own daily responsibilities, but also be able to take specific instructions for a special event and follow through to completion.<br/> <br/><br/> <br/>In addition to day-to-day activities associated with ensuring on-air signal integrity, this position provides direct leadership and is a point of escalation for their respective TOC Supervisors and management. This position advises and assists (if necessary) their direct reports and team members in resolving complex issues that arise with internal and external customers as a result of outages and failures during a particular shift. Ideally a Manager should possess in-depth operational knowledge of all Operational Support Systems (OSS) used within the TOC, have objective command of their team, build positive relationships with other teams, reviews and recommends changes to Standard Operating Procedures (SOPs) and keep senior management apprised of issues affecting their respective team.<br/> <br/><br/> <br/>In this Management position you are in a leadership role and are expected to act in a positive and professional manner at all times. You must display professionalism and act with high integrity. Managers are responsible for, but are not limited to: developing high-functioning teams and employees, recognizing and holding accountability with Teams and team members, conducting Bi-weekly one-on-ones with their direct reports, delivering regular performance feedback including performance reviews, owning difficult messages, coordinating escalations and resolving conflicts.<br/> <br/> <b>Responsibilities:</b> <br/> <br/><br/> <br/><ul><li>Direct and coordinate the activities of Supervisors and technicians with the objective of ensuring continuous operation of Dish Broadcast Center services</li><li>Build a productive, high-performance work team to deliver an advanced level of quality and service in all daily activities</li><li>Coach, mentor, and develop team to include oversight of hiring, employee onboarding and providing career development planning</li><li>Guide the professional growth of team members by writing performance appraisals and providing regular performance feedback</li><li>In collaboration with the DT Training Team, direct and monitor department formal and on-the-job training program to ensure appropriate progress</li><li>Provide expectations for team behavior and results and hold team members accountable</li><li>Recognize and reward exceptional actions and behavior above and beyond and provide disciplinary action when warranted</li><li>Administer teams timecards and absence requests in the absence of the Supervisor</li><li>Establish quality assurance guidelines, evaluate quality of teams work and recommend corrections as needed</li><li>Work in conjunction with senior leadership utilizing data or business feedback to drive performance improvements</li><li>Develop and revise policies to clarify department expectations and streamline processes</li><li>Empower employees to take responsibility for work tasks, goals and their development plans. Delegate responsibility and expect accountability and regular feedback.</li><li>Coordinate with recruiter and Human Resources to interview and hire top quality team members</li><li>Oversee the department schedule and ensure proper staffing levels for a 24x7x365 environment</li><li>Serve as an operational advisor to TOC supervisors, technicians, senior leadership as well as other departments on matters that impact TOC operations</li><li>Complete job analysis for each position and write/revise job descriptions</li><li>In conjunction with TOC leadership, develop strategic initiatives for department that contribute to achieving business objectives</li><li>Provide clear and easy to understand information to senior management on problems or issues that have occurred</li><li>Maintain a 24 hour on-call availability as needed</li><li>Ensure that all aspects of the SOPs are carried out as directed by Operations Leadership</li></ul><br/> <br/><br/> <br/><br/> <br/> <b>Basic Requirements</b> :<br/> <br/><br/> <br/><ul><li>US Citizenship or Permanent Legal Resident Status</li><li>BA/BS preferred and 4 years of progress growth or 5 years of previous management experience as a people leader</li><li>Excellent written and verbal communication skills</li><li>Excellent customer service skills</li><li>Willing to work various shifts as needed to support the TOC operations</li><li>Ability to type 20 WPM</li><li>Experience with Microsoft Office Suite and/or Google G-Suite</li></ul><br/> <br/><br/> <br/> <b>Qualifications:</b> <br/> <br/><ul><li>Ability to promote a positive environment, demonstrate respect clarify expectations, and keep commitments</li><li>Ability to comprehend system/equipment issues and provide concise statuses of system/equipment problems to senior leadership</li><li>Ability to approach work with energy and determination</li><li>Proven team player with strong customer service orientation</li><li>Adaptability: deals with ambiguity and adapts to uncertain or unexpected high-pressure situations in a calm and rational manner</li><li>Ability to read, write, analyze and interpret documents such as operation manuals, instructions and technical drawings.</li><li>Ability to write reports, business correspondence, and policies</li><li>Ability to prioritize in a fast-paced environment</li><li>Ability to work effectively with cross-functional teams</li><li>Ability to characterize and define problems, collect date, establish facts and draw meaningful conclusions</li><li>Ability to think critically, listen actively, and organize multiple sources of information to solve problems</li><li>Proven leadership skills and cohesive team player who is self-motivated and reliable to the organization; and is able to successfully interact with senior management and other departments.</li><li>Ability to foster teamwork and unity among team that allows for civil disagreement and expeditious conflict resolution, and the appreciation of diversity as well as cohesiveness, supportiveness and working effectively together to enable each employee and the department to succeed.</li></ul> Compensation: $64,800.00/Yr. - $89,400.00/Yr. <br/> <br/><br/> <br/>From versatile health perks to new career opportunities, check out our benefits on our <a href="https://careers.dish.com/life/investing-in-you/" target="\_blank" rel="noopener">careers website</a> .<br/> <br/><br/> <br/>Employment is contingent on successful completion of a pre-employment criminal background check, which may include a drug test. <br/> <br/><br/> <br/><br/> <br/> <b>Basic Requirements</b> :<br/> <br/><br/> <br/><ul><li>US Citizenship or Permanent Legal Resident Status</li><li>BA/BS preferred and 4 years of progress growth or 5 years of previous management experience as a people leader</li><li>Excellent written and verbal communication skills</li><li>Excellent customer service skills</li><li>Willing to work various shifts as needed to support the TOC operations</li><li>Ability to type 20 WPM</li><li>Experience with Microsoft Office Suite and/or Google G-Suite</li></ul><br/> <br/><br/> <br/> <b>Qualifications:</b> <br/> <br/><ul><li>Ability to promote a positive environment, demonstrate respect clarify expectations, and keep commitments</li><li>Ability to comprehend system/equipment issues and provide concise statuses of system/equipment problems to senior leadership</li><li>Ability to approach work with energy and determination</li><li>Proven team player |
| Company Name | DISH |
| Skills | [Broadcast Operations, Disciplinary Action, Advisor, Television Programming] |
| Qualification | NA |
| City | Cheyenne |
| State | Wyoming |
| Zipcode | 82001 |
| Job Opening Date | 1/7/21 |
| Job Closing Date | NA |
| Status | OPEN |
| Website Url | https://www.dish.ie |