Behavioral:

**Thank you for giving me the opportunity to be interviewed today.**

1. Tell me about yourself.

Of course!

Name

**Computer Science junior at the University of Houston’s Honors College, minoring in Mathematics and Data and Society**.

I’m interning as a Data Analyst at Smith and Associates, where I use a variety of tech stacks including Python, Javascript, SQL and MongoDB. This past summer I worked as a Software Engineering Intern, and I’m currently working on their internal L&D platform.

Over the past year, I’ve worked as an NLP/ML Researcher at Hewlett Packard Enterprise Data Science Institute, applying machine learning and NLP models to social science research.

I also serve as an advisor President of the Cougars of Data Science, leading initiatives that enhance data science skills among UH students.

2. Why should we hire you?

**unique blend of technical skills, leadership experiences, and proven multitasking abilities**

make me a strong fit for your team.

My role as the **President of Cougars of Data Science** and in **Badminton and Recreation at UH** have honed my ability to manage complex projects and lead diverse teams under pressure.

I’ve successfully **balanced a rigorous academic schedule with extensive extracurricular leadership**, managing over 500 members and orchestrating large-scale events that have significantly enhanced community engagement and operational funding.

These experiences have equipped me with **robust problem-solving skills** and an **ability to deliver high-quality outcomes**—qualities that are essential for the challenges faced in data-driven roles. My ability to adapt quickly and drive results through innovative solutions positions me as a valuable candidate for your team.

3. How would an internship at Smith & Associates help you achieve your long-term career goals?

An internship at O-I would be a crucial step in my career journey as I aim to **merge my expertise in data science with my passion for sustainable practices**. O-I's commitment to innovation in sustainable glass packaging and its substantial investment in environmental initiatives align perfectly with my long-term goals of using data science to address and solve ecological challenges. By working with the Global Sustainability and Global Engineering groups, I would have the **opportunity to refine my skills in statistical modeling and predictive analytics, applying them to impactful sustainability projects.** This experience would not only advance my technical abilities but also deepen my understanding of how data-driven solutions can foster more sustainable business practices, preparing me to be a leader in the field of environmental data science.

4. What is your greatest weakness?

A weakness I've identified in myself is that I can sometimes be **overly concise in my communication**. While this helps in being direct and efficient, I've realized that it can sometimes leave room for ambiguity, especially in detailed or complex projects. To improve, I've been **actively seeking feedback from peers** and mentors to refine how I convey information, ensuring clarity and completeness. Additionally, I'm **working on tailoring my communication style to better fit the needs of my audience**, whether it's providing more background information during presentations or more detailed explanations in project documentation.

5. What is your greatest strength?

One of my greatest strengths lies in my **adaptability**. I thrive in environments that require **quick thinking** and the ability to pivot strategies when necessary. This skill has enabled me to **effectively manage challenges** and learn new technologies or processes rapidly, ensuring that I contribute positively to my team and projects from the start. My approach combines a keen willingness to learn with a practical application of problem-solving techniques, which I believe is critical for staying ahead in fast-paced and ever-changing environments.

SITUATION

TASK  
ACTION  
RESULT

6. Tell Me About a Time When...You Overcame an Obstacle

As the President of the Cougars of Data Science at the University of Houston, I faced a significant challenge when we aimed to **significantly expand our workshop offerings and industry collaboration events**. Initially, our main obstacle was the lack of engagement from both members and potential industry partners, which threatened the success of our expanded programming.

To address this, I strategized a **multi-faceted approach to boost involvement**. First, I personally reached out to potential speakers and companies with tailored proposals highlighting mutual benefits, which significantly increased their interest in participating. Concurrently, I launched a targeted marketing campaign using our club’s social media platforms, which I enhanced by implementing data analytics to track engagement and optimize our messages.

These efforts culminated in a series of successful events that not only **doubled our active membership** but also strengthened our industry connections, providing invaluable networking opportunities for our members. Overcoming this challenge taught me the importance of proactive leadership and adaptive marketing strategies in building community engagement.

Tell me about a time you failed/made a mistake:

“In my role as President of the Cougars of Data Science, I faced a challenging situation when I decided to **overhaul our club’s workshop format** without adequately receiving feedback from the club members. My goal was to introduce more advanced topics and hands-on projects to better prepare members for real-world data science challenges. However, this shift was met with lower attendance and dissatisfaction from members who felt the new format was too advanced and not aligned with their current skills.

Realizing my mistake in not engaging the members in the decision-making process, I organized a series of feedback sessions to better understand their needs and expectations. This open dialogue helped me grasp the diverse skill levels and interests within our group. Based on this feedback, I reintroduced some beginner-friendly sessions while integrating intermediate projects with guided support, which significantly improved engagement and satisfaction.

This experience taught me the importance of member-centric leadership and the value of inclusive decision-making. It underscored that successful leadership is not just about having visionary ideas but also about ensuring those ideas resonate with and meet the needs of your team

Tell me some ways you use python, PHP and javascript at work/project

For example, I integrated PHP scripts to managecontent updates and user interactions on several websites. PHP was instrumental in handling database interactions and session management, which allowed for the creation responsive and secure user experiences.

As a Web Developer at the Honors College at UH, I utilized JavaScript to enhance front-end functionalities and user experience. One of the key projects was to implement interactive features on the college's website using JavaScript alongside HTML and CSS. These included slideshow type animations and overlapping video effects

What type of libraries have you used?  
I’ve worked witht the Lavarel libraries in PHP

React, a powerful JavaScript library for building user interfaces. React has been instrumental in developing single-page applications by allowing me to create reusable UI components that manage their state effectively. For example, in a project at the Honors College at UH, I used React to enhance the interactivity of the college's website

Notes:  
Heap is best for when you need to repeatedly look for min and max.

Min and Max heap difference

In-order tree traversal-similar to DFS

Moves like: Left, Visit, Right

Pre-order tree traversal- Print as soon as you land on a node, first print should be the root node

Pre-order: Moves like: Visit, Left, Right

Post-order: Moves like: Left, Right, Visit

Depth-first- Stack

Breadth-first - Queue

**O-I Data Science Sustainability Co-Op (Summer & Fall 2024)**

1. Tell me about yourself.

Of course!

Name

**Computer Science junior at the University of Houston’s Honors College, minoring in Mathematics and Data and Society**.

Over the past year, I’ve worked as an NLP/ML Researcher at Hewlett Packard Enterprise Data Science Institute, applying machine learning and NLP models to social science research.

II also serves as the President of the Cougars of Data Science, leading initiatives that enhance data science skills among UH students.

I am **excited to apply my technical skills and leadership experience** at a company like Smith, and I aim to contribute to innovative solutions in a new environment.

2. Why should we hire you?

**unique blend of technical skills, leadership experiences, and proven multitasking abilities**

make me a strong fit for your team.

My role as the **President of Cougars of Data Science** and in **Badminton and Recreation at UH** has honed my ability to manage complex projects and lead diverse teams under pressure.

I’ve successfully **balanced a rigorous academic schedule with extensive extracurricular leadership**, managing over 500 members and orchestrating large-scale events that have significantly enhanced community engagement and operational funding.

These experiences have equipped me with **robust problem-solving skills** and an **ability to deliver high-quality outcomes**—qualities that are essential for the challenges faced in data-driven roles. My ability to adapt quickly and drive results through innovative solutions positions me as a valuable candidate for your team.

3. How would a co-op at O-I help you achieve your long-term career goals?

An internship at O-I would be a crucial step in my career journey as I aim to **merge my expertise in data science with my passion for sustainable practices**. O-I's commitment to innovation in sustainable glass packaging and its substantial investment in environmental initiatives align perfectly with my long-term goals of using data science to address and solve ecological challenges. By working with the Global Sustainability and Global Engineering groups, I would have the **opportunity to refine my skills in statistical modeling and predictive analytics, applying them to impactful sustainability projects.** This experience would not only advance my technical abilities but also deepen my understanding of how data-driven solutions can foster more sustainable business practices, preparing me to be a leader in the field of environmental data science.

4. What is your greatest weakness?

A weakness I've identified in myself is that I can sometimes be **overly concise in my communication**. While this helps in being direct and efficient, I've realized that it can sometimes leave room for ambiguity, especially in detailed or complex projects. To improve, I've been **actively seeking feedback from peers** and mentors to refine how I convey information, ensuring clarity and completeness. Additionally, I'm **working on tailoring my communication style to better fit the needs of my audience**, whether it's providing more background information during presentations or more detailed explanations in project documentation.

5. What is your greatest strength?

One of my greatest strengths lies in my **adaptability**. I **quickly adapt to new tools and methodologies**, which would help me thrive in O-I's dynamic, innovative environment. This skill has enabled me to **effectively manage challenges** and learn new technologies or processes rapidly, ensuring that I contribute positively to my team and projects from the start. My approach combines a keen willingness to learn with a practical application of problem-solving techniques, which I believe is critical for staying ahead in fast-paced and ever-changing environments.

SITUATION

TASK  
ACTION  
RESULT

6. Tell Me About a Time When...You Overcame an Obstacle

As the President of the Cougars of Data Science at the University of Houston, I faced a significant challenge when we aimed to **significantly expand our workshop offerings and industry collaboration events**. Initially, our main obstacle was the lack of engagement from both members and potential industry partners, which threatened the success of our expanded programming.

To address this, I strategized a **multi-faceted approach to boost involvement**. First, I personally reached out to potential speakers and companies with tailored proposals highlighting mutual benefits, which significantly increased their interest in participating. Concurrently, I launched a targeted marketing campaign using our club’s social media platforms, which I enhanced by implementing data analytics to track engagement and optimize our messages.

These efforts culminated in a series of successful events that not only **doubled our active membership** but also strengthened our industry connections, providing invaluable networking opportunities for our members. Overcoming this challenge taught me the importance of proactive leadership and adaptive marketing strategies in building community engagement.

**Tell me about a time you failed/made a mistake:**

“In my role as President of the Cougars of Data Science, I faced a challenging situation when I decided to **overhaul our club’s workshop format** without adequately receiving feedback from the club members. My goal was to introduce more advanced topics and hands-on projects to better prepare members for real-world data science challenges. However, this shift was met with lower attendance and dissatisfaction from members who felt the new format was too advanced and not aligned with their current skills.

Realizing my mistake in not engaging the members in the decision-making process, I organized a series of feedback sessions to better understand their needs and expectations. This open dialogue helped me grasp the diverse skill levels and interests within our group. Based on this feedback, I reintroduced some beginner-friendly sessions while integrating intermediate projects with guided support, which significantly improved engagement and satisfaction.

This experience taught me the importance of member-centric leadership and the value of inclusive decision-making. It underscored that successful leadership is not just about having visionary ideas but also about ensuring those ideas resonate with and meet the needs of your team

Can work fully remote

From June- December

Full time in Summer, part time for rest of the year

Modernize ml models, make presentations

Design and create dashboards for ceo etc.

How would an internship at Auredia you achieve your long-term career goals?

Auredia's innovative use of deep-learning AI and blockchain for optimizing Overall Asset Effectiveness (OAE) aligns with my career goals of working on transformative technologies that not only improve business operations but also support sustainable practices.

the opportunity to work in a cutting-edge environment where I can apply my skills in Python, TensorFlow, and data analytics to practical challenges in the manufacturing sector is incredibly motivating. I am eager to contribute to and learn from a team that is pioneering the future of manufacturing with AI innovations. Working at Auredia would provide me with invaluable experience and exposure to advanced technologies, setting a strong foundation for my future career in technology and innovation.

For hashtable questions, use counter:

This tracks the frequency of every element in a hash table

sorted\_s = sorted(s) sorts an array in ascending order

This is useful when comparing 2 arrays

For Dynamic Programming Questions:

Make a decision tree- then show what we can cache to remove extra work - then write dp code

Using 2 pointers:

Use left and right to iterate through the array

1. Tell me about yourself.

Of course!

Name

**Computer Science junior at the University of Houston’s Honors College, minoring in Mathematics and Data and Society**.

Over the past year, I’ve worked as an NLP/ML Researcher at Hewlett Packard Enterprise Data Science Institute, applying machine learning and NLP models to social science research.

II also serve as the President of the Cougars of Data Science, leading initiatives that enhance data science skills among UH students.

I am **excited to apply my technical skills and leadership experience** at a company like Smith, and I aim to contribute to innovative solutions in a new environment.

2. Why should we hire you?

**unique blend of technical skills, leadership experiences, and proven multitasking abilities**

make me a strong fit for your team.

My role as the **President of Cougars of Data Science** and in **Badminton and Recreation at UH** have honed my ability to manage complex projects and lead diverse teams under pressure.

I’ve successfully **balanced a rigorous academic schedule with extensive extracurricular leadership**, managing over 500 members and orchestrating large-scale events that have significantly enhanced community engagement and operational funding.

These experiences have equipped me with **robust problem-solving skills** and an **ability to deliver high-quality outcomes**—qualities that are essential for the challenges faced in data-driven roles. My ability to adapt quickly and drive results through innovative solutions positions me as a valuable candidate for your team.

3. How would an internship at NOV help you achieve your long-term career goals?

Real-world experience, oil and gas is a booming industry

4. What is your greatest weakness?

A weakness I've identified in myself is that I can sometimes be **overly concise in my communication**. While this helps in being direct and efficient, I've realized that it can sometimes leave room for ambiguity, especially in detailed or complex projects. To improve, I've been **actively seeking feedback from peers** and mentors to refine how I convey information, ensuring clarity and completeness. Additionally, I'm **working on tailoring my communication style to better fit the needs of my audience**, whether it's providing more background information during presentations or more detailed explanations in project documentation.

5. What is your greatest strength?

One of my greatest strengths lies in my **adaptability**. I thrive in environments that require **quick thinking** and the ability to pivot strategies when necessary. This skill has enabled me to **effectively manage challenges** and learn new technologies or processes rapidly, ensuring that I contribute positively to my team and projects from the start. My approach combines a keen willingness to learn with a practical application of problem-solving techniques, which I believe is critical for staying ahead in fast-paced and ever-changing environments.

SITUATION

TASK  
ACTION  
RESULT

6. Tell Me About a Time When...You Overcame an Obstacle

As the President of the Cougars of Data Science at the University of Houston, I faced a significant challenge when we aimed to **significantly expand our workshop offerings and industry collaboration events**. Initially, our main obstacle was the lack of engagement from both members and potential industry partners, which threatened the success of our expanded programming.

To address this, I strategized a **multi-faceted approach to boost involvement**. First, I personally reached out to potential speakers and companies with tailored proposals highlighting mutual benefits, which significantly increased their interest in participating. Concurrently, I launched a targeted marketing campaign using our club’s social media platforms, which I enhanced by implementing data analytics to track engagement and optimize our messages.

These efforts culminated in a series of successful events that not only **doubled our active membership** but also strengthened our industry connections, providing invaluable networking opportunities for our members. Overcoming this challenge taught me the importance of proactive leadership and adaptive marketing strategies in building community engagement.

**Tell me about a time you failed/made a mistake:**

“In my role as President of the Cougars of Data Science, I faced a challenging situation when I decided to **overhaul our club’s workshop format** without adequately receiving feedback from the club members. My goal was to introduce more advanced topics and hands-on projects to better prepare members for real-world data science challenges. However, this shift was met with lower attendance and dissatisfaction from members who felt the new format was too advanced and not aligned with their current skills.

Realizing my mistake in not engaging the members in the decision-making process, I organized a series of feedback sessions to better understand their needs and expectations. This open dialogue helped me grasp the diverse skill levels and interests within our group. Based on this feedback, I reintroduced some beginner-friendly sessions while integrating intermediate projects with guided support, which significantly improved engagement and satisfaction.

This experience taught me the importance of member-centric leadership and the value of inclusive decision-making. It underscored that successful leadership is not just about having visionary ideas but also about ensuring those ideas resonate with and meet the needs of your team

**Questions for each project and work experience on your resume:**

1. **NLP/ML Researcher at Hewlett Packard Enterprise Data Science Institute**

Q1: Can you discuss the specific NLP and ML models you used to preprocess data for segmentation analysis and anomaly detection? What challenges did you face in tuning these models to improve data clarity and usability by 60%?

For segmentation analysis and anomaly detection, I employed a combination of supervised and unsupervised ML models, including Random Forests for anomaly detection and LSTM(Long short-term memory networks) for sequence data in segmentation analysis. The primary challenge was balancing model complexity to handle large datasets effectively, which required extensive hyperparameter tuning and validation.

This process was especially tedious because the ML models were extremely heavy and had to be opened using the Carya supercomputer.

**Q2: How did integrating the OpenAI API into your workflows enhance sentiment analysis for social science literature? Can you elaborate on the technical steps you took to achieve this integration and double efficiency?**

Integrating the OpenAI API significantly streamlined our sentiment analysis processes. I utilized the API to access GPT 3.5 which could understand human text. This was particularly beneficial for analyzing large volumes of social science literature. The integration involved setting up API calls within our existing Python frameworks and handling JSON responses for data analysis, which helped us double our workflow efficiency.

**Q3: You mentioned optimizing Python and R network analysis code for the Carya supercomputer cluster, reducing runtime by 40%. Could you describe the optimization techniques you applied? Were there any particular bottlenecks or performance issues you had to overcome?**

The optimization of network analysis code involved rewriting code for better performance and scalability. I focused on vectorizing operations in Python and parallel processing in R, using more efficient data structures.

To compute word frequencies across multiple documents, using NumPy's vectorized operations could aggregate results across the entire dataset more efficiently than iterating through each document individually.

Implementing parallel processing in R allowed you to handle larger datasets by distributing the computation across several cores of the Carya supercomputer, thereby reducing overall runtime.

Efficient data structures- using pandas dataframes in python

3. AI Denoiser for Brain MRIs Project

Q6: You developed a TensorFlow-based neural network utilizing various layers for feature extraction and dimension reduction. Could you explain the architecture of your model and why you chose specific layers like Conv2D and MaxPooling2D?

A6: The architecture of the neural network was designed specifically for image processing efficiency. Conv2D layers were crucial for extracting spatial hierarchies in images, while MaxPooling2D reduced the spatial dimensions, minimizing computational requirements. UpSampling2D was used to recover the spatial dimensions after noise reduction. This combination was effective in enhancing noise reduction while maintaining image details.

Q7: In preprocessing images for your neural network, you managed to speed up the process by 20% using Python libraries. What specific techniques or methods did you employ to achieve this improvement in preprocessing time?

A7: To speed up image preprocessing, I optimized the use of the skimage and PIL libraries by implementing batch processing techniques and utilizing multi-threading where possible. Additionally, I preprocessed images in parallel to the main training process, which significantly reduced idle times for the GPU.

5. Impact Analysis of Race and Socioeconomics on Student Discipline

Q10: You utilized a Poisson (Zero-inflated) count model with a 95% accuracy rate. Can you discuss the statistical rationale for choosing this model over others for predicting out-of-school suspensions?

A10: The Poisson (Zero-inflated) model was chosen because of the nature of the discipline data, which had many zeros (no suspensions) and over-dispersion. The zero-inflated model allowed us to separately model the occurrence of suspensions and the frequency, providing more accurate predictions and insights into the factors influencing suspensions.

Q11: What were the key findings from your regression analysis on the impact of socio-economic and racial factors on student discipline? How did these findings inform the policy recommendations you developed?

The key findings from this predictive analysis were that students from economically backward backgrounds and students of hispanic and black had a higher chance of receiving one out of school suspension (OSS)

ML RECAP

Linear Regression = Least Squares

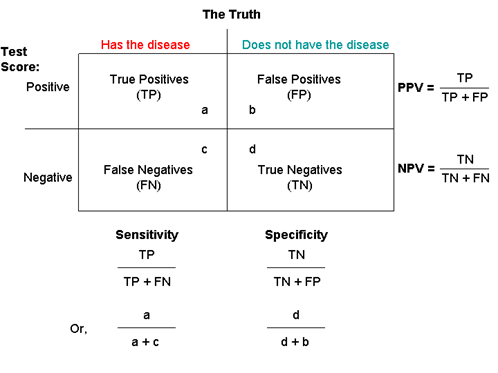
ML models deal with training and testing data to predict something

Cross-Validation: Compares diff ML models and checks how well they fit with the data

CV takes diff 80-20 splits each time, and summaries them at the end

If we divide data into 4 blocks(3 training, 1 testing), it’s called 4-fold cross validation

LOOCV- Each sample is considered a block



Sensitivity- How many peeps with disease were correctly identified

Specificity- How many peeps without the disease were correctly identified

Variance = Difference in fits between datasets

Receiver Operator Charactersitc(ROC) shows the easiest way to summarise the data

True Positive Rate = Sensitivity

False Positive Rate = 1 - Specificity

AUC = Area Under Curve of ROC graph

Trees are pruned by replacing 2 leaves by a leaf that is the average of the two.

This is done using sum-squared

Bagging is used to reduce variance and help avoid overfitting. It is particularly effective for high-variance, low-bias models (like decision trees).

Example: random forests are an ensemble of decision trees

Boosting is used to reduce both bias and variance in a model. It is effective for improving weak learners and is generally considered a powerful way to increase model accuracy.

Example: AdaBoost (Adaptive Boosting)

If you had this type of data or system, how would you build a model?  
  
Example Question 1: Predictive Maintenance for Equipment

Question: If you were given sensor data from oil drilling equipment, how would you build a model to predict equipment failure?

I would try and follow these broad steps to set up a prediction model

**Data Understanding and Preparation:** First, I would examine the sensor data to understand the types of sensors involved (e.g., temperature, pressure, vibration) and the frequency of data collection. Key steps would include handling missing values, removing outliers, and feature engineering to create meaningful attributes (like moving averages or change rates).

**Exploratory Data Analysis (EDA):** Conduct EDA to find patterns, trends, and anomalies in the data. This would help in identifying which features are most relevant to equipment failure.

**Model Selection:** **For predictive maintenance, time-series forecasting** models could be useful. I might start with simpler models like logistic regression to set a baseline and then explore more complex models like Random Forests for better accuracy.

**Model Training and Validation:** Split the data into training and test sets, ensuring that the split reflects different operational periods rather than random sampling. This is crucial for time-series data to prevent leakage. Cross-validation would be used to tune parameters and avoid overfitting.

**Deployment and Monitoring**: Once the model is trained and validated, I would deploy it in a test environment to monitor its predictions against actual equipment performance. Continuous monitoring would be essential to adjust the model as new data comes in or conditions change depending on seasons and wear-and-tear of machines