Parent Company – Official Name and Address: *	Motif Research, Inc., DBA Artifact 4949 E Sagebrush Ln Eagle Mountain, Utah 84005-7220 970 658-6909
Parent Company DUNS # (Look up DUNS# here):	117415607
Name and address of your organization: *	Motif Research, Inc., DBA Artifact 4949 E Sagebrush Ln Eagle Mountain, Utah 84005-7220 970 658-6909

Briefly describe your organization: *

We're a revenue generating SaaS startup focused on structuring and extracting insights from large quantities of unstructured data. Specifically, we warehouse enterprises' consumer interactions that occur through mediums like Intercom chats, Salesforce records, sales call transcripts, email records, and ERP system records. After warehousing, we organize the data, generate actionable topics covering what their customers are saying, summarize that information using Machine Learning generated natural language summaries, and make it available to our clients through our SaaS solution.

Website:	http://artifact.io

Describe the capstone problem: *

We pull data from our client's systems at various cadences to provide insights into unstructured data. Across our 30+ clients and 50+ data connectors, we retrieve data from multiple systems like Salesforce, Intercom, Zendesk, Apple app store reviews, Qualtrics survey results, etc. Often clients have custom implementations of these systems, and our connectors are customized to handle it. This leads to a wide variety of connectors and data we pull in.

The variety brings challenges. Quality checks on our data feed and various transformation steps in our Machine Learning pipeline are rudimentary. For example, suppose the client has changed their system. In that case, the volume of daily data we pull might drop, and that's not noticeable immediately. So an ongoing challenge for us is to audit the connectors and various steps in the data pipeline and implement automated data quality monitoring steps.

During the capstone, we hope to expose the students to the problem of auditing data connectors and monitoring data quality in large-scale information systems. They would apply and evaluate the use of statistical and ML techniques to identify data volume and distribution deviations proactively. They would then deploy these ML-powered monitoring solutions to production systems within our data pipeline.

Objectives of the project: *

- Students would access our Google Cloud Platform and perform all the analyses within that environment.

- Students will lead an early phase of consulting style audit where they interview the Artifact team and understand the various steps within our data pipeline and connectors. Part of this would require students to review SQL / Python code, understand its inner workings, and analyze data quality by sampling data from various steps in the pipeline.

- Students would then identify the highest value steps in the pipeline where we can apply automated statistical or ML-powered data quality monitoring. They would focus on identifying data volume shifts and data distribution shifts.

- Students would attempt to implement a few methods by working with the Artifact team and integrating them with the existing data pipeline.

- Students would document the learnings around various statistical and ML methods available for monitoring data distribution shifts.

What is the business value of the project to your organization? *	- The project would improve transparency in our data pipeline and help proactively monitor data quality. This directly translates to a better customer experience and decreases the number of bugs/tickets raised by our customers. - With pro-active monitoring of data distribution shifts, the ML team can identify clients whose data is changing and might need a model fine-tuning, thereby reducing customer churn due to poor results.
Will your organization provide data to the student team? *	Yes
What type of data will you provide for the students? Check all that apply: *	• Confidential/Proprietary Data
Confirm that the data you share will contain NO Personally Identifiable	• Confirmed

Information (PII): *	
Confidential/Proprietary Data: *	The data is confidential to our clients. It's unstructured text data retrieved from their business systems like Zendesk - which tracks tickets and customer support chats.
Public Data: *	NA
Recruiting Interests:	• Do you currently post open positions for your organization at Heinz College? • Are you interested in hiring students as a result of the capstone project?
Program preferences: *	• Master of Information Systems Management (MISM) • Business Intelligence and Data Analytics
Skills/technologies required: *	SQL, Statistics and Data Mining are the core skills. Knowledge of Python is an added benefit.
Preferred Semester:	Fall: August to December
Capstone Lead Contact Name: *	Trey Davis
Job Title: *	Co-founder & Operations
Email: *	trey@artifact.io
Phone Number: *	(208) 569-1504
Additional Capstone Company Contact Name:	Nate Sanders
Job Title:	CEO
Email:	nate@artifact.io
Phone Number:	(970) 658-6909
Legal Signatory ***Who will be your legal signatory on the CMU Capstone Agreement?*** **Name:** *	Nate Sanders
Job Title: *	CEO
Email: *	nate@artifact.io