

Cohort 31 - Group 4									
Vanessa Gleason, Alistair Marsden, Olivier Rochaix, Eduard Stalmakov									
Capstone Project									
Project Management Plan									
Milestone	Task	Assigned to	Estimated time (hrs)	Estimated Start	Estimated Finish	Task Status	Name	Number of Tasks	Task Hours
Capstone Outline							Eduard		
	Create Github repository for collaboration	Eduard	0.5	9/16/22	9/16/22	Complete	Vanessa		
	Start and strucutre a README file for GitHub	Eduard	1.5	9/16/22	9/16/22	Complete	Alistair		
	Decide on exploratory questions	All	6	9/20/22	9/20/22	Complete	Olivier		
	Stored on GitHub as "ExploratoryQuestions.pdf"	Eduard	0.5	9/20/22	9/20/22	Complete			
	Find datasets [user playlists, song attributes, and user demographics]	All	12	9/20/22	9/20/22	Complete			
	Find a dataset to link census data to audio streaming market	All	6	9/22/22	9/23/22	IP			
	Find census datasets [Age, Income, Computer Access]	All	6	9/21/22	9/21/22	IP			
	Research machine learning algorithm	All	12	9/21/22	9/21/22	Complete			
	Create elevator pitch formal summary	All	2	9/20/22	9/20/22	Complete			
	Complete initial PMP	All	4	9/22/22	9/22/22	Complete			
	Stored on GitHub as "ProjectManagementPlan.pdf"	Vanessa	0.5	9/22/22	9/22/22	Complete			
	Create ER diagram	Eduard	2	9/21/22	9/21/22	Complete			
	Data platform napkin drawing	Olivier	4	9/20/22	9/20/22	Complete			
	Create ETL document structure	Vanessa	1	9/22/22	9/22/22	IP			
	Stored on GitHub as "RepeatableETLReport.pdf"	Vanessa	0.5	9/22/22	9/22/22	IP			
Access									
	Create SQL database	Eduard	0.5	9/20/22	9/20/22	Complete			
	Create group container	Eduard	0.5	9/20/22	9/20/22	Complete			
	Create Kafka topic	Eduard	0.5	9/20/22	9/20/22	Complete			
	Create a Data Factory	Eduard	1	9/20/22	9/20/22	Complete			
Submit on Github									
Produce and Consume Kafka Data using Data Factory									
	Create producer in DataBricks	Eduard	1	9/22/22	9/22/22	IP			
	Read last.fm songs into topic with producer	Eduard	1	9/22/22	9/22/22	IP			
	Set up data factory pipeline from data set to the producer in Data Factory	Eduard	1	9/22/22	9/22/22	IP			
	Create consumer in DataBricks to collect messages from the Kafka server	Eduard	1	9/22/22	9/22/22	IP			
Reading Kafka Files from Data Lake Into the Database									
	Create one or more tables for the finished messages you are consuming	Alistair	1	9/22/22	9/22/22	IP			
	Set up data factory pipeline from from the producer to the consumer	Eduard	1	9/22/22	9/22/22	IP			
	Export the data to the group database	Eduard	1	9/22/22	9/22/22	IP			
Building and Exporting an ML Model									
	EDA on data in database	Olivier	2	9/22/22	9/22/22	IP			
	Build PowerBI test ML model	Vanessa	3	9/22/22	9/23/22	IP			
	Use Joblib to export the model in Python	Alistair	3	9/22/22	9/23/22	IP			
	Save your exported model to an Azure Blob	Alistair	1	9/23/22	9/23/22	IP			
Build a Power BI Report and Dashboard									
	Load the model into Power BI using Python	Olivier	1	9/23/22	9/23/22	IP			
	Test that ML model works in Power BI	Olivier	1	9/23/22	9/23/22	IP			
Diagram and Document Your Cloud Servers									
	Detail and document the set of tools used / planned for use	Olivier	1	9/22/22	9/22/22	IP			
	Stored on GitHub as "DataPlatform.pdf"	Olivier	0.5	9/22/22	9/22/22	IP			
Submit on Github									
Napkin Drawings and Feedback for Visualizations									
	Create Napkin Drawing								

Cohort 31 - Group 4									
Vanessa Gleason, Alistair Marsden, Olivier Rochaix, Eduard Stalmakov									
Capstone Project									
Project Management Plan									
Milestone	Task	Assigned to	Estimated time (hrs)	Estimated Start	Estimated Finish	Task Status	Name	Number of Tasks	Task Hours
	Stored on GitHub as "VisualizationsNapkinsAndFeedback.pdf"								
	Get two napkin Drawing feedback								
	Document Feedback								
	Stored on GitHub as "DashboardNapkinsAndFeedback.pdf"								
	Add feedback to Dashboard								
Submit on Github									
	Create data factory all necessary pipelines	Eduard		9/20/22					
	use the databrick to put kafka data into an azure data lake	Eduard		9/20/22					
	use azure data factory to move data from the azure data lake to azure SQL								
Cloud ETL must include:	A network diagram or service diagram of data pipelines								
	Identify which services are going where								
	Identify how things are getting moved								
Machine Learning Model	Machine learning research	all	5						
	Building the machine learning model								
	Refining the machine learning model for higher accuracy								
Dashboard	Choose Dash or Power BI for visuals	Alistair	3	9/21/22	9/21/22				
	investigate PowerBI	Alistair	3	9/21/22	9/21/22				
	Create Dashboard								
	include the Dash dashboard or Power BI Report files in a folder in your repo called "dashboard"								
	Add Data Source to Dashboard								
	Choose color theme	all	0.25						
	assign/make visuals (barchart, graph, heat map, pie, etc)	Vanessa	3						
	assign/make visuals (barchart, graph, heat map, pie, etc)	Alistair	3						
	assign/make visuals (barchart, graph, heat map, pie, etc)	Eduard	3						
	assign/make visuals (barchart, graph, heat map, pie, etc)	Olivier	3						
	If there is any code involved for the deliverables: make a github folder called "code" and store all code there								
	Complete Dashboard								
Submit on Github									
Project Technical Report	Start Report								
	Exploratory Questions								
	Introduction								
	Research Presentation								
	Machine Learning								
	Conclusion								
	Review Report and Complete								
	Stored on Github as ProjectTechnicalReport.pdf								
Submit on Github									
Presentation Slides	Create Google Slides								
	Create Presentation								
	Presenting Plan								
	Stored on GitHub as "CapstonePresentationSlides.pdf"								
Submit on Github									
Presentation Dry Run	Final Review of All Work								

Cohort 31 - Group 4									
Vanessa Gleason, Alistair Marsden, Olivier Rochaix, Eduard Stalmakov									
Capstone Project									
Project Management Plan									
Milestone	Task	Assigned to	Estimated time (hrs)	Estimated Start	Estimated Finish	Task Status	Name	Number of Tasks	Task Hours
	Practice Presenting								
	Present for grading				10/6/22				
	Present to Employers				10/7/22				