

Cohort 31 - Group 4							
Vanessa Gleason, Alistair Marsden, Olivier Rochaix, Eduard Stalmakov							
Capstone Project							
Project Management Plan							
X	ID broad topic - ecommerce > music streaming service						
IP	Identify 3 data sets (plus the 1 census dataset)						
	all teammates explore and find a potential concept						
X	create GitHub repository - Eduard						
	https://github.com/EduardStalmakov/Capstone-Project						
IP	come up with 5-10 questions about data in your industry						
	stored on GitHub as " ExploratoryQuestions.pdf "						
	one question has to be a predictions, classifications, or find clusters of some sort ML						
	formalize questions						
	we'll go through the questions a few times during week of 9/19						
	Create elevator pitch formal summary to sen			Due 9/20/22			
IP	create project management plan			Due 9/23/22			
	stored on GitHub as " ProjectManagementPlan.pdf "						
	Setting up platform for project			Projected 9/19-9-23			
	Regarding Kafka and Cloud ETL:						
	must have Kafka Producer Stream						
	must configure and consume topics						
	capture the kafka stream in a databrick						
	use the databrick to put kafka data into an azure data lake						
	use azure data factory to move data from the azure data lake to azure SQL						
	Specifically to Cloud ETL you must include:						
	A network diagram or service diagram of data pipelines						
	Identify which services are going where						
	Identify how things are getting moved						
	create napkin drawings and feedback for the visualizations						
	save all screenshots and include them in a document						
	stored on GitHub as "VisualizationsNapkinsAndFeedback.pdf"						
	create napkin drawings and feedback for the dashboard						
	save all screenshots and include them in a document						
	stored on GitHub as "DashboardNapkinsAndFeedback.pdf"						

	GitHub repository must include:							
	5-10 questions about data stored on GitHub as "ExploratoryQuestions.pdf"							