IE 7275: Project Guidelines

Objective	To demonstrate the applications of data mining principles and processes on a practical problem.
Group Effort	Students work in groups of two on the project
Description	The project is intended to give students a hands-on experience of the entire data analytics process, including business problem definition, solution design, data selection and/or collection, data processing, data exploration, data reduction, transformation, and variable selection, model building, algorithm implementation, and predictive performance evaluation, visualization and reporting. Students can choose a dataset of their own choice from any data source except `https://www.kaggle.com/`

Sample Data Sources	**UCI Machine Learning Repository**
	`http://archive.ics.uci.edu/ml/index.php`
	US Open Data Project
	`https://www.data.gov/`
	Financial Data
	`https://www.quandl.com/`
	Awesome Public Datasets
	`https://github.com/awesomedata/awesome-public-datasets`
	Datasets Subreddit
	`https://www.reddit.com/r/datasets/`
	Google BigQuery Public Datasets
	`https://cloud.google.com/bigquery/public-data/`
	100 plus free data sources
	`https://www.columnfivemedia.com/100-best-free-data-sources-infographic`

Cover page	Add a cover page to every milestone you submit.
Grading	The project is designed to test your ability to apply your fundamental understanding of the material to a practical problem unlike a well-structured homework problem. Your interim reports serve the purpose of documents' steady progress on their projects. Interim reports are not graded. If students need feedback on their projects at any point in the semester, they are encouraged to make an appointment with the

instructor or TA to discuss. The projects are evaluated at the end of the semester based on the following criteria:

Project selection and problem definition 10%
Data collection 10%
Data exploration, visualization and processing 10%
Dimension reduction and variable selection 10%
Model exploration and selection 10%
Model performance evaluation 10%
Performance visualization 10%
Study progress through the semester 10%
Project presentation 10%
Report organization/writing/clarity 10%