ML Project ideas : Sheet1

	A	B	C	D	E	F	G
1		**ML Algorithm**	**Problem Statement**	**Project Idea**	**Public Dataset**	**Predictor Variable**	**Outcome Variable**
2	Supervised	Simple Linear Regression	How does square footage affect house prices?	Predicting house prices based on square footage	https://www.kaggle.com/c/boston-housing	Continuous	Continuous
3		Multiple Linear Regression	How do study hours and attendance affect student performance?	Predicting student performance based on study hours and attendance	https://archive.ics.uci.edu/ml/datasets/Student+Performance	Continuous (2 or more predictors)	Continuous
4		Logistic Regression	Will a customer buy a product based on their characteristics?	Predicting if a customer will buy a product (yes/no)	https://www.kaggle.com/datasets/carrie1/ecommerce-data	Continuous	Binary
5		Decision Trees	How do leaf measurements predict the species of a plant?	Classifying species of plants based on leaf measurements	https://archive.ics.uci.edu/ml/datasets/Iris	Continuous or Categorical	Continuous or Categorical
6		Random Forest	How can we detect fraud in credit card transactions?	Fraud detection in credit card transactions	https://www.kaggle.com/mlg-ulb/creditcardfraud	Continuous or Categorical (multiple trees)	Continuous or Categorical
7		Support Vector Machines	Can we classify images of cats and dogs based on their features?	Image classification (e.g.identifying cats vs dogs)	https://www.cs.toronto.edu/~kriz/cifar.html	Continuous or Categorical	Binary or Continuous
8		K-Nearest Neighbors	How does the proximity to stores affect house prices?	Recommender system for movies based on user ratings	https://grouplens.org/datasets/movielens/	Continuous or Categorical	Continuous or Categorical
9		Gradient Boosting Machines	How can we predict customer churn in a subscription service based on usage patterns?	Predicting customer churn in a subscription service	https://www.kaggle.com/blastchar/telco-customer-churn	Continuous or Categorical	Continuous or Categorical
10		Naive Bayes	How can we classify emails as spam or not spam based on their content?	Email spam detection	https://spamassassin.apache.org/publiccorpus/	Categorical	Categorical
11		Neural Networks	How can we recognize handwritten digits?	Handwriting recognition (e.g.MNIST dataset)	http://yann.lecun.com/exdb/mnist/	Continuous or Categorical	Continuous or Categorical
12	Unsupervised	K-Means Clustering	How can we segment customers based on their purchasing behavior?	Customer segmentation for targeted marketing	https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial	Continuous	Clusters
13		Hierarchical Clustering	How can we group similar articles based on text content?	Grouping similar articles based on text content	http://qwone.com/~jason/20Newsgroups/	Continuous	Clusters
14		Principal Component Analysis (PCA)	How can we reduce the number of features in a dataset while preserving variance?	Reducing features in a high-dimensional dataset for visualization	https://archive.ics.uci.edu/ml/datasets/wine	Continuous	Principal Components (Continuous)
15		Independent Component Analysis (ICA)	How can we separate mixed audio signals into their original sources?	Separating audio signals from a mixed recording	http://www.cis.hut.fi/projects/ica/cocktail/cocktail_en.cgi	Continuous	Independent Components (Continuous)
16		Apriori Algorithm	What product combinations are commonly bought together?	Market basket analysis for retail stores	https://archive.ics.uci.edu/ml/datasets/online+retail	Categorical	Association Rules
17		DBSCAN	How can we identify clusters of crime incidents in geographic data?	Identifying geographic clusters of crime incidents	https://www.kaggle.com/shivamb/real-time-crime	Continuous	Clusters
18		t-Distributed Stochastic Neighbor Embedding (t-SNE)	How can we visualize high-dimensional genetic data?	Visualizing high-dimensional genetic data	https://www.internationalgenome.org/data-portal/data-collection/phase-3	Continuous	Lower-Dimensional Embedding (Continuous)