MSA 8050 SCALABLE DATA ANALYTICS
PROJECT PROPOSAL - TEAM 3
INDUSTRY ANALYSIS USING YELP REVIEWS

➢ **MOTIVATION:**

Yelp is a platform which publishes crowd-sourced reviews about businesses, serving as a reliable link between businesses and potential customers.

In the past few years, we have witnessed an unprecedented change in every consumer facing industry. These uncertainties affecting the businesses arise due to a host of external factors such as pandemic, looming fear of a war, supply-chain disruption, just to name a few. Amidst this, we aim to identify such industries that serve the customers unwaveringly. To do so we analyze the businesses and industries that thrived during these times and present insights into their locations, check-in times, operational days etc. through the paradigm of top performing reviewers.

Additionally, we would also like to decipher the correspondence among the reviews posted by all the top reviewers through the vocabulary used in their respective reviews.

➢ **PROPOSED CONTRIBUTION:**

- ➢ **Shruti Shivaji Lanke** (002583158) - Data pre-processing and initial data analysis by leveraging RDDs and Spark Data-frames.
- ➢ **Suvaleena Paul** (002638763) - Insight gathering and analysis using RDDs and Spark data frames
- ➢ **Pranshu Srivastav** (002602585) - Machine Learning pipeline and model fitting using NLP (Natural Language Processing).
- ➢ **Ankita Mitra** (002591022) - Power Point Presentation preparation.
- ➢ **Tania Haldar** (002253341) - Project Report preparation.

➢ **PROJECT FLOW:**

Industry analysis using Yelp reviews