# Machine Learning Project Proposal:

## H-1B Visa Petitions 2011-2016

For the CS 584 course, the dataset that was chosen to further analyze in this project is called “H-1B Visa Petitions 2011-2016”.

As data scientists with little experience in the financial field, we found it would be a good opportunity to learn a bit more while applying our knowledge in statistics to this area of expertise.

The chosen dataset includes is an Excel file with all H-1B Visa Petitions from 2011 to 2016.

This dataset was interesting to us because in the future it is very likely that we are going to fill a H-1B Visa petition and we thought it would be useful to us and a lot people to have a model like this to know before hand if they are likely to get their Visa.

The attributes in the dataset are:

1. **Case Status**: If the petition was certified, denied or withdrawn.
2. **Employer name**: Name of the company.
3. **Soc name**:
4. **Job title**: What was the position held by the requester.
5. **Full time position**: If position was full time or not.
6. **Prevailing wage**: salary.
7. **Year**: What year was the petition made.
8. **Worksite**: Where was the person working.
9. **Lon**: Longitude coordinate of the worksite.
10. **Lat**: Latitude coordinate of the worksite

The initial idea is to create a classification model that could tell us if a person, based on the attributes, is more likely to be given or denied the Visa.

We are going to create several models using different techniques and compare their performance to find the best one.