A Project Proposal on

*Population density shifts & unemployment*

By: Team Member’s Name:

**Tehauger Hauger**

**Sneha Godbole**

**Brett Steele**

Table of Contents

[1. Problem Statement 3](#_Toc446007330)

[**1.1** **Introduction** 3](#_Toc446007331)

[**1.2** **Team** 3](#_Toc446007332)

[**1.3** **Planned Roles** 3](#_Toc446007333)

[**1.4** **Project Type** 3](#_Toc446007334)

[**1.5** **Tools and Technologies** 3](#_Toc446007335)

[2. Deliverables 4](#_Toc446007336)

[**2.1** **Artifacts** 4](#_Toc446007337)

[**2.2** **Documentation** 4](#_Toc446007338)

[**2.3** **Schedule** 4](#_Toc446007339)

[3. Acknowledgement 4](#_Toc446007340)

# **Problem Statement**

## **Introduction**

Given US Census Bureau dataset for 2010, We will pre-process the data and store it in Apache Hadoop and using data mining algorithms we will identify the population density shifts in ZCTA’s (Zip Code Tabulation Areas) and how it relates to the unemployment rates in the area. We will also create dashboards for displaying concentrations using different visualization techniques.

* 1. **Team**
* Thomas Hauger : **IU Username**: tehauger; **GitHub**:  [tehauger@indiana.edu](mailto:tehauger@indiana.edu); futuresystems.org: tehauger
* Sneha Godbole: **IU Username**: snegodbo; **GitHub:** [snegodbo@umail.iu.edu](mailto:snegodbo@umail.iu.edu) ; futuresystems.org: sneha21
* Brett Steele: **IU Username**: steeleb, **GitHub**: [steeleb@indiana.edu](mailto:steeleb@indiana.edu); futuresystems.org: brttstl
  1. **Planned Roles**
* Obtain and format data: Thomas Hauger
* Map/Reduce Functions: Sneha Godbole
* Coding: Brett Steele
* Visualization Dashboard: Sneha Godbole and Thomas Hauger
* Statistical Analysis: Brett Steele
  1. **Project Type**
* Basic and Deployment

## **Tools** **and Technologies**

Following tools and technologies will be used for implementation of the solution:

1. Processing and Analyzing Dataset
   1. Apache Hadoop
   2. ArcGIS
   3. MongoDB
   4. Java
   5. Python/R
2. Creating and Visualizing Dashboards
   1. D3.js/Tableau
3. Documentation
   1. Microsoft Office
4. Compute Resource
   1. OpenStack in futuresystems
5. System Requirements

* Virtual Machine instances
* Ubuntu
* Storage

# **Deliverables**

Following artifacts / reports would be submitted as a part of deliverables of the project –

## **Artifacts**

1. US Census Bureau Dataset being used.
2. Statistical tables and code.
3. A D3.js library for implementing and visualizing dashboards.

## **Documentation**

Following sections would be added to this document –

1. Implementation Details
2. Dashboard Snapshots

## **Schedule**

* Class Week Ending 3/18/2016: Initial Planning and Submit Proposal
* Class Week Ending 3/25/2016: Discussion and Brainstorming
* Class Week Ending 4/1/2016: Presentation
* Class Week Ending 4/8/2016: Refine Data
* Class Week Ending 4/15/2016: Build Systems
* Class Week Ending 4/22/2016: Develop statistics and visualizations
* Class Week Ending 4/29/2016: Submit Project

# **Acknowledgement**

* Project idea obtained from list of potential projects presented in class.
* Chose idea after not being able to obtain data from Infochimps concerning Uber.