

# W200 Python Fundamentals for Data Science, UC Berkeley MIDS

## Project 2 Draft Proposal

Group Members: C. King, V. Meausoone, V. Nalini, E. Petrov

As at July 26, 2019

Github repo: [Project2 King Meausoone Nalini Petrov](#)

## Evaluating the Quality and Cost of Palliative and Hospice Care in the U.S.

### Context

The U.S. Census Bureau<sup>1</sup> (USCB) has flagged the year 2030 as “an important demographic turning point in U.S. history” as this is the year when the USCB expects approximately 20% of U.S. residents to be older than age 65, as compared to 13% in 2010. This rapid demographic change is expected to transform the U.S. economy and social fabric. While this demographic change has many social and economic implications, one of the most critical to understand is the stress it will place on our already fragile healthcare system, particularly as it relates to end-of-life care (or EoLC).

EoLC generally involves two critical health care services: (1) palliative and (2) hospice care. Palliative care encompasses specialized medical services for people living with a chronic, terminal or serious illness. Palliative care focuses on alleviating symptoms and supporting the overall well-being of both the patient and caregivers. In contrast, hospice care focuses on supporting the patient’s quality of life once treatment options are exhausted and the illness is deemed incurable. Typically, hospice care is initiated within the last 6 months of the patient’s life and while it may incorporate palliative components, hospice care does not involve a full-time medical team.

The U.S. Government administers two separate healthcare programs (i.e., Medicare<sup>2</sup> and Medicaid<sup>3</sup>) that influence the delivery and cost coverage of palliative and hospice care. Palliative care is a covered medical expense under Medicaid, whereas hospice care relies upon Medicare’s hospice benefit.

The purpose of our project is to conduct an exploratory data analysis (EDA) of key healthcare variables tracked for palliative and hospice services in the U.S. in order to understand the utilization, quality, and cost of the current services provided.

### Key Research Questions

- How extensive are palliative and hospice care services in the United States? How are these services distributed across states and counties?
- What are the (1) quality, (2) utilization and (3) cost characteristics of palliative and hospice care services delivered in the U.S.? Is there significant geographic variation?
- How do patient satisfaction levels differ between wealthy and poor regions?

### Hypotheses

Given that palliative services are only partially covered by health care programs in the U.S., we expect that (1) income levels will directly impact the quality and cost metrics, and (2) the demand, quality and cost structure will vary at the state/county level.

### Scope of Work

The table below summarizes the core competencies our project will demonstrate.

---

<sup>1</sup> <https://www.census.gov/newsroom/press-releases/2018/cb18-41-population-projections.html>

<sup>2</sup> Medicare is an insurance program administered by the federal government that provides health coverage for patients over the age of 65 (or under 65 for those with a disability) regardless of the patient’s income.

<sup>3</sup> Medicaid is a state and federal assistance program that serves low-income individuals regardless of age.

Competency	Scope								
<b>Data Preparation</b>	<p><u>Data Sourcing</u></p> <ul style="list-style-type: none"><li>Listed below are three data sets we will use in our analyses. We selected these data sets based on our research questions and a preliminary review of the data. The first two data sources will comprise our main data set. The remaining data sources will be used for data enrichment of the main data set.<ol style="list-style-type: none"><li>Medicare synthetic files<sup>4</sup> for Inpatient, Outpatient and Beneficiary Claims</li><li>Hospice compare data set<sup>5</sup></li><li>U.S. Census data set<sup>6</sup></li><li>Mapping of FIPS state and county to ZIP code<sup>7</sup></li></ol></li><li>Ingest data from multiple data sources in different data formats<ul style="list-style-type: none"><li>Create ingest “classes” to import data</li><li>Use methods to implement different ingestion techniques</li></ul></li></ul> <p><u>Data Cleaning/Wrangling</u></p> <ul style="list-style-type: none"><li>Review the data and conduct any needed data cleansing for the selected variables of interest.</li><li>Apply filters to select individuals eligible for hospice care</li><li>Identify needed variables, or derive new variables from raw data if required<ul style="list-style-type: none"><li>Based on the raw synthetic Medicare data:<ul style="list-style-type: none"><li>Number of hospitalizations per person</li><li>Total cost of care per person</li></ul></li><li>Based on census data:<ul style="list-style-type: none"><li>Mean or median income by geographical unit (i.e. county)</li></ul></li><li>Based on Hospice compare:<ul style="list-style-type: none"><li>Quality of care indicators.</li></ul></li></ul></li><li>Harmonize the data sets and establish links (i.e., bridge between state/county identifiers in main data set to zip codes in the census data set)</li><li>Create a single analytical dataset to be analyzed with pandas</li></ul>								
<b>Data Analysis</b>	<p><u>Review and Evaluate the Data</u></p> <ul style="list-style-type: none"><li>Import Python packages for the purpose of analysis (i.e., demonstrating functional programming)</li><li>Demonstrate our knowledge of different Python data types and methods for handling data from the disparate source data (e.g., for loops, while loops, etc.).</li><li>Conduct EDA of key explanatory variables</li><li>Refine selection of explanatory variables and refine EDA</li></ul>								
<b>Data Visualization</b>	<p><u>Define Visualization Class</u></p> <ul style="list-style-type: none"><li>Create a visualization class that implements methods to handle different graphical outputs (e.g., histogram, bar graph, etc.).</li></ul> <p><u>Create Visualizations</u></p> <ul style="list-style-type: none"><li>Demonstrate we can use functional programming to import the visualization class and get it to work (e.g., Seaborn, Folium, etc.).</li><li>Use built in Python package (Matplotlib) to create visualizations.</li><li>We intend to depict the variables listed below using the stated visualizations:</li></ul> <table><tr><th>Variable</th><th>Intended Depiction</th></tr><tr><td>Quality of Care</td><td>Heat map across geographies (overlaid on a map of the U.S.)</td></tr><tr><td>Cost</td><td>Funnel chart based on geography</td></tr><tr><td>Utilization</td><td>Heat map across geographies (overlaid on a map of the U.S.)</td></tr></table>	Variable	Intended Depiction	Quality of Care	Heat map across geographies (overlaid on a map of the U.S.)	Cost	Funnel chart based on geography	Utilization	Heat map across geographies (overlaid on a map of the U.S.)
Variable	Intended Depiction								
Quality of Care	Heat map across geographies (overlaid on a map of the U.S.)								
Cost	Funnel chart based on geography								
Utilization	Heat map across geographies (overlaid on a map of the U.S.)								

<sup>4</sup> [https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE\\_Syn\\_PUF.html](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html)

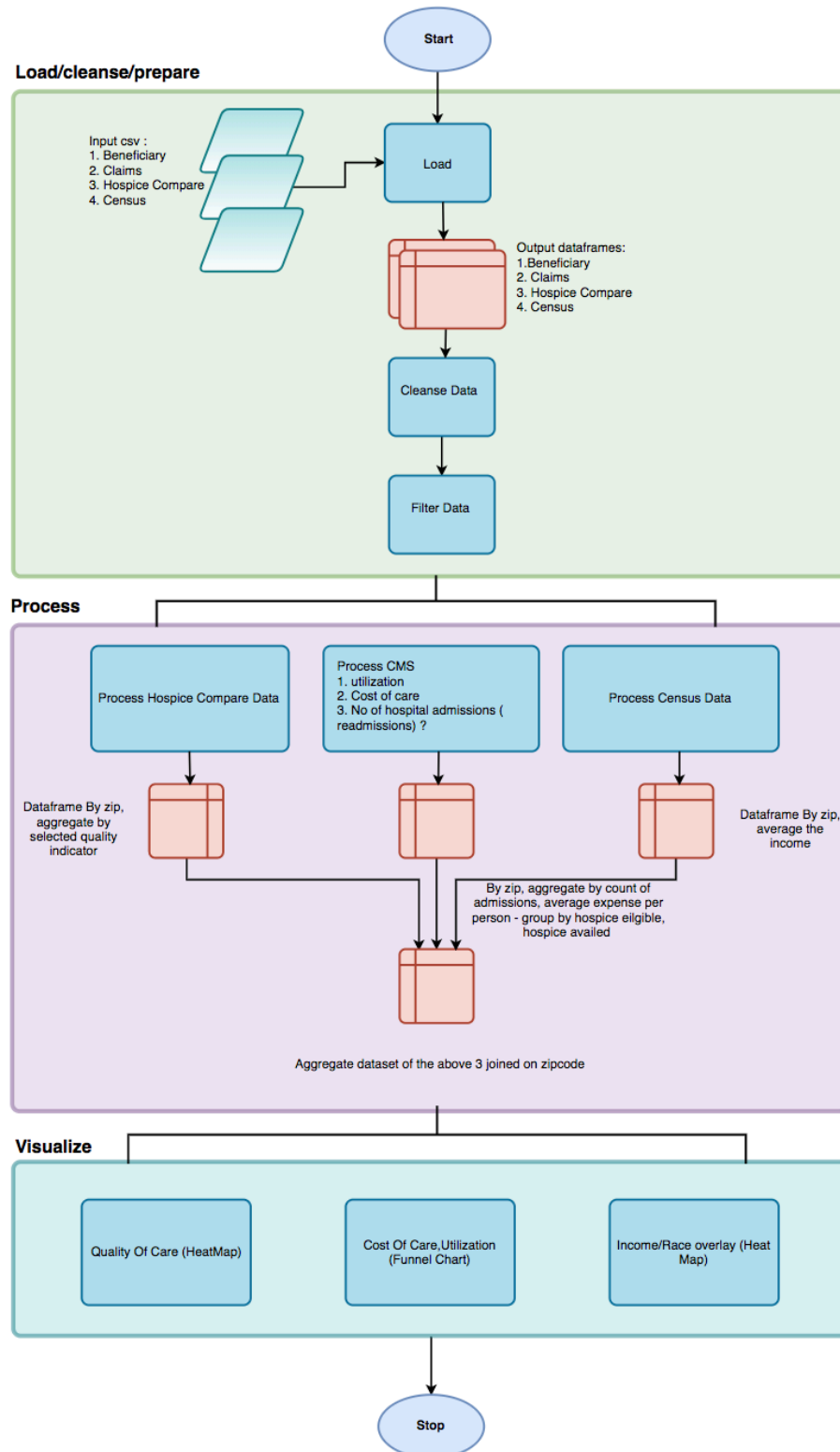
<sup>5</sup> <https://data.medicare.gov/data/hospice-compare>

<sup>6</sup> <https://www.census.gov/programs-surveys/metro-micro/data/datasets.html>

<sup>7</sup> <https://www.kaggle.com/danofer/zipcodes-county-fips-crosswalk>

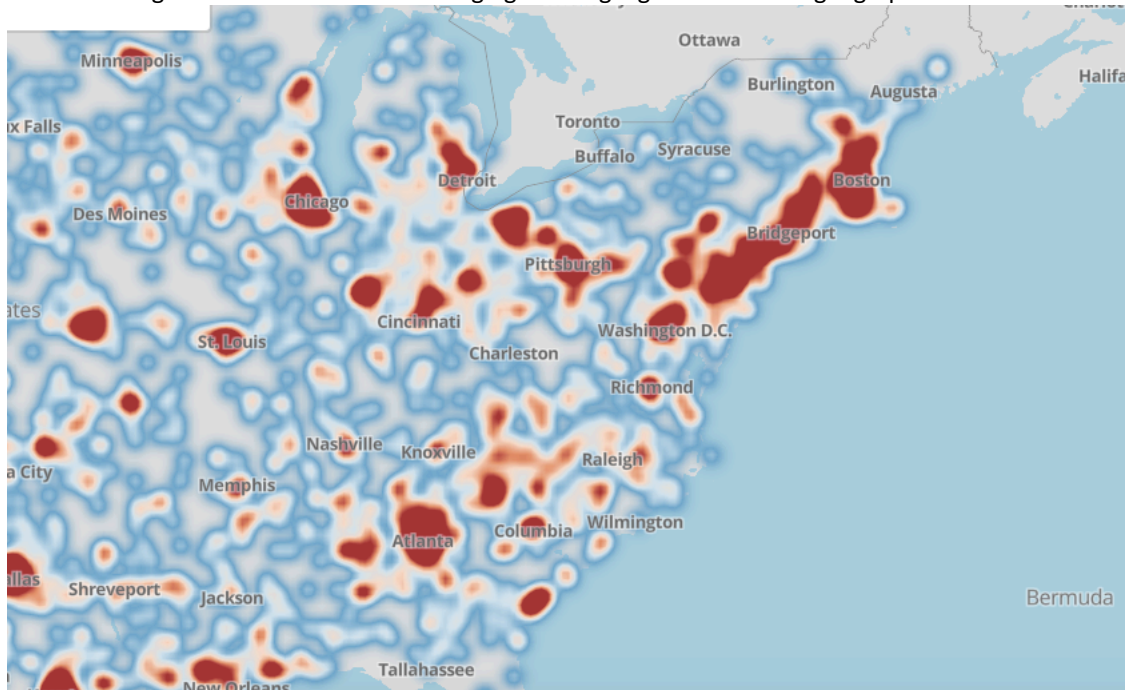
## Preliminary Architecture

### Workflow: Palliative Care Utilization and Cost by State/County in US



### Example of expected visuals:

- Distributions of cost of care by State in the US. This will provide a sense of the outliers.
- A heat map showing the variation in each measure of interest: utilization, costs and quality. Heatmaps will allow understanding where measures are converging to being high for the same geographical area.



Source: Hospice Compare Data, Medicare.

- A table comparing two US States – one which relies on hospice care and another that does not - on more granular measures of utilization, cost of care and quality of care. We will be able to derive how these differences in hospice care delivery translate in terms of our measures of interest.

### Final Report:

The final report of this project will include the following elements:

- An introduction: Similar to our proposal, we will explain why this project is compelling to us, both from a domain-specific perspective, and in demonstrating our knowledge of Python acquired during this course.
- A methods section: This method will describe how we ingested the data, how we curated the variables to create our measures of interest, and how we linked/overlaid disparate data sources. It will also explain how our cohort of hospice care patients was identified, and what our assumptions were to generate it.

This section will also cover how we approached the project from a coding perspective. Specifically, it will detail our rationale for using functional programming and object-oriented programming in distinct parts of the project and why. Finally, it will cover our approach for creating our visuals.

- A results section: This section will display our charts and tables, such as the one included in the section above “example of expected visuals”. It will cover our analytical observations. We will also discuss limitations in our ability to draw insights, if applicable, and suggest ways to enhance this work going forward.
- Conclusion: This section will provide a short summary of this project, highlighting successes and challenges.