1. Get to know your data. Briefly describe the attributes that will help you in the following tasks (in bullet points)

Data contains information about populations of churches per county of the United States of America. It consists of 3075 rows ( counties ) and 235 columns.

Columns consist of:

* CaseID : an id for the entry
* CNAME: County name
* STCODE: The State code
* CCODE: County code
* TOTPOP: The total population of the County
* TOTMEMB: The number members of any church in the County
* TOTCHUR: The number of Churches in the County
* Columns that end with ‘\_M’ eg : AOG\_M, denoting the number of members of Church AOG in the County.
* Columns that end with ‘\_C’ eg : AOG\_C, denoting the number of Churches AOG in the County.

TOTCHUR and Columns that end with ‘\_C’ will be discarded as only the population will be used to determine the location of the cross religion center.

2) Summarize the data to help understand the overall picture of religious groups over the US.

A barplot of total population per Church(titles of Churches are omitted for reporting reasons):

A graph with blue lines

AI-generated content may be incorrect.

As we can see there is a big number of churches with low polulation. Few have extremely high population compaired to others, let's see which are

those churches.

CATH\_M 29688058

METH\_M 8790025

SBC\_M 8121069

JEWS\_M 5112024

EPISC\_M 2544320

Catholic, Methodist, Southern Baptists, Jews and Protestant Episcopal

Church are the 5 most populated churches.

3) Which are the counties with the highest per-person ratio of Orthodox

Christian members?

To calculate this we sum columns which refer to Christian orthodox Churches

(ARAPO\_M, BEOC\_M only are kept as GRKAD\_M, ACROC\_M have zero population reported) and divide it by the TOTPOP of each County.

Results are:

A screenshot of a cell phone

AI-generated content may be incorrect.

Cherokee.AL, Madison.IL and Fresno.CA have the highest per-person ratio of Orthodox Christian members.

4) Find the 3 most extreme counties with respect to the distribution of their churches across religions.

To find the 3 most extreme counties with respect to the distribution of their churches across religions we will use LOF on church population distribution vectors.

LOF is based on a concept of a local density, where locality is given by *k* nearest neighbors, whose distance is used to estimate the density. By comparing the local density of an object to the local densities of its neighbors, one can identify regions of similar density, and points that have a substantially lower density than their neighbors. These are considered to be outliers.

In our case every county is plotted in a 107 dimensional space in which every dimension is the proportion of its TOTEMB (the members of any church) that belong in the columns Church. Neighbors for LOF calculation set to 50.

The counties with the three biggest LOFs are :

239 Hinsdale, CO

1813 Kings, NY

1955 Billings, ND

5) Where would you create a cross-religion center of discussion between religions to maximize its impact? Support the proposal based on data analysis results.

A Cross-Religion Center should be located in an area with highreligiousdiversity to ensure representation and inclusivity. It must be geographicallycentral for accessibility to a large portion of the population and minimize transportation.

As a first step we will build hypothetical per Religion centers with the use of weighted k-Means with 1 centroid where weights will be the population of every county for the specified religion . After that we will again perform weighted k-Means with 1 centroid to the per Religion centers , this time the weights will be the total Religion population. The location of the resulting centroid will be recommended location for the cross religion center, every church will be weighted according to its population count and positioning on the US map.

To achieve that we downloaded coordinates for every county FIPS code from kaggehub (path :‘alejopaullier/usa-counties-coordinates/ cfips\_location.csv’)

Results for per religion centroids are :

A screenshot of a computer

AI-generated content may be incorrect.

Results for the coordinates of the Cross-Religion Center are :



An interactive Dash figure is created to visualize the results . In the first plot one can see a heatmap of the population of every church along with the centroid derived from the first weighted k-Means. Churches can be changed through the scrollbar .In the second plot one can see every centroid , with size depending on the total population of each church, along with a red star which is the proposed location of the cross religious center.

A map of the united states

AI-generated content may be incorrect.

A map of the united states with a red star

AI-generated content may be incorrect.