Assignment 1
Cloud Computing
193RSEG-176-1DL, Fall 2, 2019
Instructor: Ari Davidow
Submitted by: Daniel Joyner & Kevin Markvenas

GROUP C
MARKVENAS – MASSACHUSETTS ANALYIC RESOURCE OF KNOWLEDGE VALUATION FROM EPIDEMIOLOGICAL NETWORK ASSESSMENTS DATABASE

November 5, 2019

OVERVIEW

1. Project Background and Description

MARKVENAS (Massachusetts Analytic Resource of Knowledge Valuation from Epidemiological Network Assessments) is a not for profit organization whose mission is to vastly improve the Massachusetts health profile by providing a community accessible knowledge discovery tool.

Problem:

Currently, more and more public health focused entities and researchers are coming to the realization that some of the biggest determinants of health outcomes are not behavioral or biological, but environmental[1,2]. Unfortunately, efforts to provide searchable databases to better understand the multiple factors which do impact public health are either difficult to navigate, only give statistical reports, or are only readable by subject matter experts[3,4,5]. It is our belief that epidemiologic factors impact community health as much as biological and behavioral risk factors, and, with better insight, individuals and community leaders can make better decisions that improve health outcomes. Our tool combines research data from a wide number of publicly available databases and presents location specific information. What differentiates our tool from other similar tools like the Center for Disease Control Interactive Atlas of Heart Disease and Stroke (http://www.cdc.gov/dhdsp/maps/atlas) or the City Health Dashboard (https://www.cityhealthdashboard.com), is that our tool goes beyond reporting data and provides contextual information about certain metrics and their relation to health outcomes, thus servicing a need for educating the public. Stakeholders for this service will primarily be lay people, but it may also serve as a resource for health and healthcare system professionals as well as hospitals to put mandatory community health needs assessment feedback in context.

2. Project Scope

The pilot for the project acts as a proof of concept. The database is accessible through the URL http://www.markvenas.com (to test the site, go to the bottom of the document). The user will be met with a landing page with a search box (below) seeking Zip Code or City input (as of now, only data for the state of Massachusetts is available).

---

[1] Bhatt, J. (2018, May 16). Your Zip Code, Your Health: AHA News. The American Hospital Association. Retrieved from https://www.aha.org/news/insights-and-analysis/2018-05-16-your-zip-code-your-health.

[2] Yitshak-Sade, M., James, P., Kloog, I., Hart, J. E., Schwartz, J. D., Laden, F., … Zanobetti, A. (2019). Neighborhood Greenness Attenuates the Adverse Effect of PM2.5 on Cardiovascular Mortality in Neighborhoods of Lower Socioeconomic Status. *International journal of environmental research and public health, 16(5), 814. doi:10.3390/ijerph16050814*

[3] https://www.cityhealthdashboard.com/

[4] http://chnafinder.hret.org/

[5] https://www.cdc.gov/dhdsp/maps/atlas/

At this stage of development, a user can enter in a zip code, city or town in the state of Massachusetts. The site and will return location specific information on **three** health related metrics:

1. *Cardiovascular disease death rate (per 100K by county)*

2. *Life Expectancy*

3. *Poverty (% of population in total poverty by county)*

The user will also receive their City, Zip Code, and County information, as well as the population of the specific city. This will be followed by short information about each metric and their relation to health outcomes (e.g. what factors may affect life expectancy in a particular area, how income correlates with life expectancy, or how cardiovascular disease death rates are linked to location).

Finally, the user will be given links to further information and resources to assist in taking action toward improving their health outcomes by taking measures for personal improvement or by taking actions in their community toward a healthier environment (e.g. advocating for better particular waste management [deployable in a future iteration of the site]).

The scope of this project is only for the state of Massachusetts. However, we are utilizing Amazon Web Services (AWS) with the intention for scalability to the national level in the future.

Customers/ Audience:

The database will be oriented toward public use. Visitors to the site we expect to be primarily lay people and non-professionals. However, information gained through the resource can be utilized by professionals and organizations.

Stakeholders:

Stakeholders for the site will include grant funding entities who will seek deliverables of a positive trend in community health outcomes.

Tracking impact:

Every three years hospitals produce community health needs assessments. We can correlate changes in community needs with high-risk metrics called out through our database, to check for an inverse correlation. We can also use analytics to track site traffic, and conduct surveys and have other reporting mechanisms to assess the manner and context our resource is being referenced in the community.

3. High-Level Requirements - AWS

The database will make use of Amazon Web Services (AWS)

This data driven solution will be benefitted by using services made available through Amazon Web Services. The steps to building the database are as follows:

Data was downloaded from publicly available sources and loaded into a SQL database which is designed to read and correlate relative information (in this case correlating census tract IDs across Zipcode, Life Expectancy, Poverty and CVD data).

The interface for the searchable database will be a Website that we setup and maintain. In AWS we created an RDS DB instance of a SQL server in the Ohio East region. It is named "cloudtest".
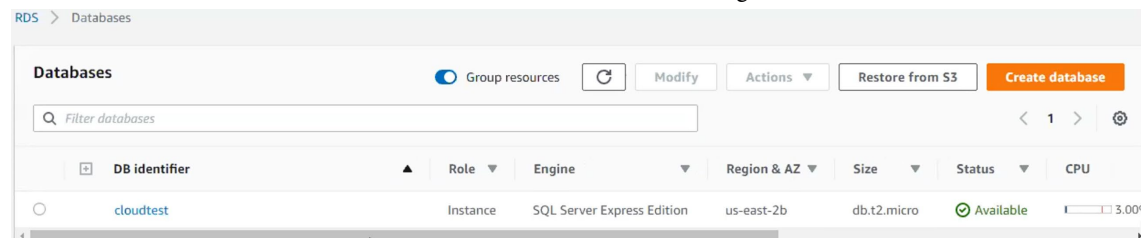
Access to the "cloudtest" instance is controlled by a Virtual Private Cloud (VPC) security group and firewall rules (see Figure 1). The website will be hosted on an Elastic Cloud Computing (EC2) server. Right now, our instance of the SQL server is running on a template provided by AWS on a virtual network (Microsoft Windows Server 2012 R2 Base), utilizing .NET development environment. We launched an EC2 server with Microsoft Server 2012 R2 and Internet Information Systems (IIS). We also included the

database on AWS so that the site is all Amazon driven. All the image files are being stored in an S3 storage bucket (not shown). The public website accesses the database through the VPC instance (Figure 1).
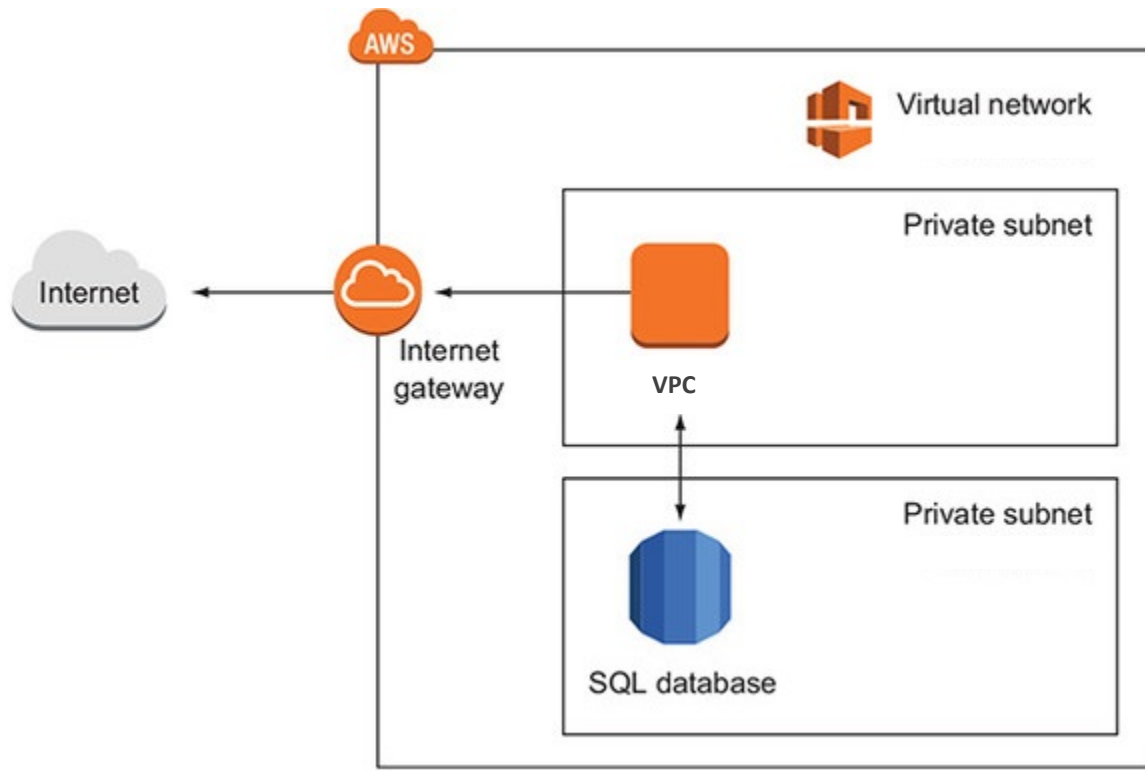

Figure 1 (Adapted from Wittig,2018)[6]

4. Benefits of using AWS (**Wittig, 2018**)[6]

- By Hosting the website on an EC2 server, we have the capability of launching as many instances of our application as we desire on any operating system. This will provide the flexibility for scalability in the future.

- By utilizing the Virtual Private Cloud service, we can specify what IP addresses can access the server, making it more secure.

- AWS makes estimating the resources that are needed to launch, maintain, and grow a project such as this easier.

- Working with AWS also is more cost effective, because we will not have to build and maintain the physical infrastructure to host this service.

- Working with AWS we also do not have to worry about backups, as AWS is highly reliable and fault tolerant

- AWS allows for flexible scalability. This project is designed for growth into other markets. AWS will more easily allow for this by allowing us to multiply instances of our application in a wide number of regions where the requisite data is available. We also have plans on increasing functionality and

---

[6] Wittig, Andreas & Michael. (2018). Amazon Web Services in Action, 2e. Manning Publications Co.

customization of variable selection, which will require growing our database. This is much easier to do using virtual systems than using offline databases.

- Also, as we expand into different markets, AWS will allow us to work on the resource from anywhere, as well as store and access feedback data, which will be in the future directions for the project.

5. Stakeholders/ Deliverables

Entities with an interest in public health:

- The American Hospital Association of any of its Professional Membership Groups[7]

- The Robert Wood Johnson Foundation

- The Centers for Disease Control

Individuals working in the health space and with an interest in public health:

- Public Health Consultants

- Physicians

- Community Researchers interested in Public Health

These aforementioned groups will find use for this database. Deliverables for these entities include:

- A single source to explore data from disparate research entities related to public health.

- In the future, feedback trends for tracking variances in imminent health dangers by community

Primarily the database is for public use for users to:

- Look at the state of health metrics for their location (e.g. life expectancy, CVD death rate, ect)

- Get more information about the health impacts of obscurely related metrics like poverty rate or particulate concentration (future)

- Get resources for more information or assistance

6. Specific Exclusions from Scope

As this is a pilot project and a proof of concept, we limited the scope of the information and customization of the searchable database. As it stands, a user can enter in their zip code city or town within the state of Massachusetts and get returned data on Life Expectancy, Cardiovascular disease (CVD) death rate, and Poverty percentage.

Limitations:

- Only county level data exists for metrics:

---

[7] https://www.aha.org/pmgs

- o   CVD death rate

- o   % Poverty

- The search function is not customizable with respect to what data is returned to the user.

- *The database is limited to Massachusetts data*

7. **Test the site**

Click below to be redirected to the testing page or follow the link to:

http://www.markvenas.com/

8. **Future Directions**

- ✓ Future development of this project will see it grow into a wider market, eventually servicing the entire United States.

- ✓ Users will have the option of selecting to have the tool select for them a list of the largest health threats with respect to their area (e.g. pollution, CVD, smoking, etc), or be able to select desired metrics and have returned the current statistics on those metrics for the selected area.

- ✓ Additionally, the feedback and resources given will be customized based on the selection and feedback. Also, resources will be constantly checked for dead links and only the top, most vetted resources will be made available.

APPROVAL AND AUTHORITY TO PROCEED

We approve the project as described above, and authorize the team to proceed.

Name	Title	Date
Daniel Joyner	Team Leader	11/05/2019
Kevin Markevas	Second Team Leader	11/05/2019
Ari Davidow	CEO	

Approved By	Date	Approved By	Date
Ari Davidow			