# **Teacher Notes for Diabetes Health Campaign**

## **Motivation and Essential Understandings**

Pima Indians living in Mexico have a diabetes prevalence of about 8%, whereas those who have emigrated to the USA, where the lifestyle is more sedentary and access to energy dense (fatty) food is easier, have a diabetes prevalence of about 50% respectively. The N.I.H. believes that if it can understand why the Pima are so obese it can better understand obesity in the rest of us.

* Is Type 2 Diabetes Mellitus caused by a person’s environment (i.e., from obesity caused by overnutrition) or by their genetic makeup?

## **Context and Dataset**

Students in Life & Health Sciences will interpret visualizations using data from the **Pima Indian** dataset based on a population of women who were at least 21 years old, of Pima Indian heritage and living near Phoenix, Arizona. This population was tested for diabetes according to World Health Organization criteria. Students will use this information to suggest a public health campaign.

## **Learning Objectives**

Students will be able to:

## Describe the Pima Indian population in terms of historical health data

## Explore correlations associated with diabetes

## Use data visualizations to explain predictors for diabetes

## Apply analysis in the context of creating a public health plan

## **Data Science Concepts and Skills**

1. Summary statistics
2. Exploratory data analysis; Static data visualization
3. Data wrangling
4. Hypothesis testing
5. Predictive modeling with Decision tree; Logistic Regression

## **Students**

This lesson is for late undergraduate students. Students should be familiar with statistical concepts, basic data visualizations, and have worked in Excel.

## **Time to Teach this Lesson**

This lesson can be taught in 2 sessions using a worksheet

**First Week**: 1-hour prep, 2-hour class session; guided data analysis and review of Statistics

**Second Week**: 1-hour class session to review decision trees; guided & independent work to suggest elements of a public health campaign

## **Lesson Materials**

You will find all the lesson materials in the GenAI GitHub repository. The R notebook is not necessary to teach this lesson but is available to those who wish to teach more hands-on Data Science.

|  |  |  |
| --- | --- | --- |
| **Materials** | **File** | **Description** |
| Lecture | Lecture\_Overview\_Diabetes\_Health\_Campaign\_2020.pptx | PPT lecture on Pima Indians |
| Handout 1 | Handout\_NDReport\_ Diabetes\_Health\_Campaign\_2020.pdf | National Diabetes Stats Report |
| Handout 2 | Handout\_PimaPardox\_ Diabetes\_Health\_Campaign\_2020.pdf | Pima Paradox article New Yorker, 1998 |
| Worksheet | Worksheet\_Diabetes\_Health\_Campaign\_2020.docx | Lesson worksheet for health campaign |
| Video 1 | Intro to AI.mp4 | Patrick Martin talks AI (20 min) |
| Video 2 | Intro to ML.mp4 | Patrick Martin talks ML (20 min) |
| Dataset |  | Pima Indian Diabetes dataset |
| Data dictionary |  | pdf of data dictionary explaining the column headings (data fields) in the datasets |
| R notebook | JupyterNotebook\_Jupyter\_ Diabetes\_Health\_Campaign\_2020.pdf | R notebook |
| R notebook pdf | JupyterNotebook\_pdf\_ Diabetes\_Health\_Campaign\_2020.pdf | pdf of annotated R notebook; Health Analysis of Diabetes among Pima Indians |
| Storyboard | Lesson\_Storyboard\_ Diabetes\_Health\_Campaign\_2020.docx | Lesson planner storyboard |

## **Teaching Strategies**

* Oral quiz in class to check students’ understanding of diabetes: terms, measures, issues
* Review concepts in Statistics such as descriptive stats, correlation, and hypothesis testing
* Pose **challenge questions** for engagement and allow students to interpret visualizations and hypothesize. Students may have difficulty limiting inferences to within the scope of the dataset, so discuss over-hypothesizing beyond the data.
* Discuss the difference between Inferential vs. Predictive modeling

## **Lesson Narrative**

|  |
| --- |
| **Module 0: Pre-lesson** |

Ask students to read foundational articles on Pima Indians and their struggle against diabetes. Provide link to CDC site describing diabetes.

Review concepts of descriptive Statistics, correlation, and hypothesis testing.

Ask students to view AI & ML videos to prepare for a discussion on the concepts of AI/ML modeling

|  |
| --- |
| **Module 1: Dataset and Exploratory Data Analysis** |

This lesson narrative follows the lesson worksheet. Interpretive questions are posed for each step.

Illustrate **Descriptive Stats**: erroneous data; distributions; correlations between features

|  |  |  |
| --- | --- | --- |
| **Module 2: Explain Predictive Modeling** | | |

Present and demonstrate two models for prediction.

* **Logistic Regression:** Confusion matrix; binary dependent variable (diabetes, yes/no?)
* **Decision tree**: Demonstrate over-correlation – binary variable for diabetes prediction; Confusion matrix: Precision, Accuracy, Recall

|  |
| --- |
| **ASSIGNMENT:** Ask students about feature importance and how to identify predictors. Students will discuss and suggest how might these findings help create an effective health campaign. |

|  |
| --- |
| **Module 3: Close Out** |

Post-assessment questions.

* What are some of the limitations of using this dataset?
* How does context help frame your interpretation?