Analysis of Characteristics of Top Dog Breeds

Alyson M Dunkle

Western Governors University

**Table of Contents**

[A. Project Overview 3](#_Toc58328659)

[A1. Research Question or Organizational Need 3](#_Toc58328660)

[A2. Context and Background 3](#_Toc58328661)

[A3. Summary of Published Works 3](#_Toc58328662)

[A3a. Relation of Published Works to Project 3](#_Toc58328663)

[A4. Summary of Data Analytics Solution 3](#_Toc58328664)

[A5. Benefit to Organization and Decision-Making Process 3](#_Toc58328665)

[B. Data Analytics Plan 3](#_Toc58328666)

[B1. Goals, Objectives, and Deliverables 3](#_Toc58328667)

[B2. Scope of Project 3](#_Toc58328668)

[B3. Standard Methodology 4](#_Toc58328669)

[B4. Timeline and Milestones 4](#_Toc58328670)

[B5. Resources and Costs 4](#_Toc58328671)

[B6. Criteria for Success 4](#_Toc58328672)

[C. Design of Data Analytics Solution 4](#_Toc58328673)

[C1. Hypothesis 4](#_Toc58328674)

[C2. Analytical Method 5](#_Toc58328675)

[C2a. Justification of Analytical Method 5](#_Toc58328676)

[C3. Tools and Environments of Solution 5](#_Toc58328677)

[C4. Methods and Metrics to Evaluate Statistical Significance 5](#_Toc58328678)

[C4a. Justification Of Methods and Metrics 5](#_Toc58328679)

[C5. Practical Significance 5](#_Toc58328680)

[C6. Visual Communication 6](#_Toc58328681)

[D. Description of Datasets 6](#_Toc58328682)

[D1. Source of Data 6](#_Toc58328683)

[D2. Appropriateness of Dataset 6](#_Toc58328684)

[D3. Data Collection Methods 6](#_Toc58328685)

[D4. Data Quality 6](#_Toc58328686)

[D5. Data Governance, Privacy and Security, Ethical, Legal, and Regulatory Compliance 6](#_Toc58328687)

[D5a. Precautions 6](#_Toc58328688)

[E. Sources 6](#_Toc58328689)

# Project Overview

# A1. Research Question or Organizational Need

State the question/need that your project will address. Summarize its significance and the context within which it is relevant.

# A2. Context and Background

More completely describe the context and background for your project. This is the scenario within which the question/need exists. Offer enough information so that the reader can understand why the question/need exists and is important to resolve.

# A3. Summary of Published Works

In this section summarize the content of the 3 articles related to the project.

# A3a. Relation of Published Works to Project

In this sub-section, discuss how each published work relates to the project and its planning.

# A4. Summary of Data Analytics Solution

Restate the business problem/need and the extent to which the proposed project will address it. The connection between the problem and solution should be clear.

# A5. Benefit to Organization and Decision-Making Process

An organization does not exist in a vacuum. Who are the stakeholders/users/decision-makers? How will the project’s process (execution) and outcomes help them?

# Data Analytics Plan

# B1. Goals, Objectives, and Deliverables

State the project’s goal or goals. Each goal must have at least 2 objectives to mark the process for achieving the goal, and each objective must terminate in at least one deliverable. A single goal is OK. Feel free to include a simple table to illustrate the relationship.

The goal of this project is to analyze a dataset of top dog breeds to discover the top three dog breeds for which a breeder could choose from to start breeding.

The objectives for this goal are:

* + - Determine eight characteristics, four negative and four positive, that can be used for comparison.
      * The deliverable for this objective is a list of the eight characteristics which can be used.
    - Determine the top three dog breeds of each of the seven dog breed categories.
      * The deliverable for this objective would be a list of the top three dog breeds of each category.
    - Determine the overall top three dog breeds.
* The deliverable for this objective would be a final list of the overall top three dog breeds.

# B2. Scope of Project

Clearly state what your solution includes, and what it does not include.

# B3. Standard Methodology

Discuss your project management methodology and how you plan to proceed in each of its phases. Example methodologies are Waterfall, ADDIE, SDLC, CRISP-DM, SEMMA and KDD.

# SEMMA – Sample, Explore, Modify, Model, and Assess

# B4. Timeline and Milestones

Present a table showing for each milestone its projected start and end dates, and its projected duration:

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone** | **Projected Start Date** | **Projected End Date** | **Duration (days/hours)** |
| Cleaned data in Jupyter Notebook | 7/5/2022 | 7/6/2022 | 8 hours |
| Created graphs in Tableau | 7/6/2022 | 7/11/2022 | 8 hours |
| Complied data of top dog breeds in Excel | 7/6/2022 | 7/11/2022 | 4 hours |
|  |  |  |  |

# B5. Resources and Costs

Present a table showing personnel, technology, and infrastructure costs. Add any needed descriptive detail below the table.

# B6. Criteria for Success

How will you measure the success and effectiveness of outcomes? Present at least 3 criteria for success, including how you will collect the data for each criterion and what constitutes success. Summarize your discussion in a table:

|  |  |  |
| --- | --- | --- |
| **Criterion/Metric** | **Required Data** | **Cut Score for Success** |
|  |  |  |
|  |  |  |
|  |  |  |

# Design of Data Analytics Solution

# C1. Hypothesis

Clearly state the hypothesis/need identified in A1.

# C2. Analytical Method

Clearly identify the descriptive or predictive method to be used.

Examples of acceptable descriptive methods:

* K-means clustering
* Hierarchical clustering
* Other clustering methods (max distance, min distance, etc.)
* PCA (all variables need to be numeric)
* MCA (all variables need to be factorized)
* FAMD
* Logistic regression with interpretations of estimated coefficients
* Multi-linear regression with interpretations of estimated coefficients

Examples of appropriate non-descriptive methods:

* Logistic regression
* Decision trees
* Random forest
* Neural network
* Multi-linear regression (if they provide a strong justification – like ease of interpretability)

# C2a. Justification of Analytical Method

Discuss why you chose this method. What makes it appropriate for this data and the question it is intended to address.

# C3. Tools and Environments of Solution

Identify the tool to be used for data extraction and why it is selected over other tools.

# C4. Methods and Metrics to Evaluate Statistical Significance

What are the methods and metrics you will use in the tool, and how is statistical significance determined?

# C4a. Justification Of Methods and Metrics

Justify the appropriateness of these metrics and methods as the correct ones to use.

# C5. Practical Significance

Statistical significance indicates the existence of an effect, while practical significance deals with the magnitude of the effect. How practical is this result in the context of decision-making? Present examples of how this result will matter in practical use.

# C6. Visual Communication

What graphical representations will you use to communicate the findings? Why use these over others that may be chosen?

# Description of Datasets

# D1. Source of Data

List the dataset(s) used for analysis. List them in **E. Sources** and cite them here, all in APA format.

# D2. Appropriateness of Dataset

Justify the choice of this dataset, or these datasets, as appropriate for addressing the project goal(s).

# D3. Data Collection Methods

Offer a discussion of how data was collected, including the steps taken to insure accuracy and quality. What are the advantages and disadvantages of collecting data using this methodology? Also discuss any challenges in the collection process and how they were overcome.

# D4. Data Quality

Discuss issues with the quality of the data and the need for interventions. Missing data, outliers, formatting issues, dirty data, and other data quality categories could be discussed.

# D5. Data Governance, Privacy and Security, Ethical, Legal, and Regulatory Compliance

These considerations could relate to the problem/need, data sensitivity and nature, the industry within which the problem/need exists, etc.

# D5a. Precautions

Discuss guidelines related to data preparation, analysis, storage, access, and dissemination.

# Sources

Michael Worboys, Julie-Marie Strange, & Neil Pemberton. (2018). *The Invention of the Modern Dog: Breed and Blood in Victorian Britain*. Johns Hopkins University Press.

Ghirlanda, S., Acerbi, A., Herzog, H., & Serpell, J. A. (2013). Fashion vs. Function in Cultural Evolution: The Case of Dog Breed Popularity. *PLoS ONE*, *8*(9), 1–6. https://doi.org/10.1371/journal.pone.0074770