

# Data Management and Data Analytics Capstone Topic Approval Form

## Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information.

**Student Name:**

**Student ID:**

**Capstone Project Name:** Analysis of US Electric Car Registrations and Renewable Energy Production by State

**Project Topic:** This project will examine state-level data on electric car adoption in conjunction with data on the sources of electrical energy production in those states.

**Research Question:** Is there a correlation between the level of green/renewable energy sources used to produce electrical power in a state and the rate of electric car adoption in that state?

**Hypothesis:** The use of green/renewable energy sources will have a positive correlation with electric car adoption in a given US state.

**Context:** The (fictional) Green Ad Council wants to encourage electric vehicle adoption in the United States. The group has previously studied the effects of state subsidies/incentives and is looking for additional criteria that might indicate which states would be good targets for marketing and customer education campaigns. The Green Ad Council believes that states which source higher percentages of their electrical power from green/renewable sources could be strong candidates for additional outreach campaigns -- particularly those states that have made progress in cleaner energy production but have not yet seen a boost in electric vehicle registrations. Many customers are unaware of the sources used to produce their electrical power. The Green Ad Council feels this information could be useful in their public communication strategy and also as an indicator of which states have a demonstrated commitment to clean energy going forward.

**Data:** I will need to collect state-level data on the breakdown of electrical power sources along with data on electrical versus total car registrations.

I will combine data from 3 existing US Government datasets:

- Car registrations by state (2018) - US Dept. of Transportation (<https://www.fhwa.dot.gov/policyinformation/statistics/2018/xls/mv1.xlsx>)
- Electric vehicle registrations by state (2018) - US Department of Energy (<https://afdc.energy.gov/data/10962>)



- Annual energy generation by state (1990 - 2019) - US Energy Information Agency ([https://www.eia.gov/electricity/data/state/annual\\_generation\\_state.xls](https://www.eia.gov/electricity/data/state/annual_generation_state.xls))

The datasets are made publicly available by the US Federal Government. I will not use any restricted, private, or proprietary data.

**Data Gathering:** I will download tabular data from the US Government websites. I will cleanse and combine the data as needed.

**Data Analytics Tools and Techniques:** I will apply descriptive statistical methods to produce aggregations and derived values for the state power generation data. For example, I will produce percentages of total power generation by energy source for each state. I will then perform exploratory data analysis and linear regression to gauge the possible correlation between clean energy source usage and electric car adoption in each state. I will use standard Python and/or R libraries and custom code to assist in data acquisition, cleansing, manipulation, and analysis.

**Justification of Tools/Techniques:** I will need to filter and summarize the power generation data -- tasks which use common descriptive analytics techniques. To explore the potential correlations between the power generation and electric car registration data, I will need to perform bivariate and potentially multivariate analysis. This will involve producing scatter plots and performing linear regression.



**Application Type, if applicable (select one):**

- ☐ Mobile  
☐ Web  
☒ Stand-alone

**Programming/Development Language(s), if applicable:** Python, R

**Operating System(s)/Platform(s), if applicable:** N/A (cross-platform)

**Database Management System, if applicable:** N/A

**Project Outcomes:** The project will produce a clean dataset from the source US Government data, including summarized and derived values. It will include the code used in the data analysis. The project will produce a report summarizing state-by-state clean energy production, electric car adoption, and an assessment of any correlations between them.

**Projected Project End Date:** 6/30/2021

**Sources:** 1) NRDC - Renewable Energy: The Clean Facts (<https://www.nrdc.org/stories/renewable-energy-clean-facts>). 2) Forbes - The Five Factors Driving the Mass Adoption of Electric Vehicles (<https://www.forbes.com/sites/enriquedans/2021/01/24/the-five-factors-driving-the-mass-adoption-of-electricvehicles>) 3) New York Time: How Does Your State Make Electricity? (<https://www.nytimes.com/interactive/2020/10/28/climate/how-electricity-generation-changed-in-your-state-election.html>)

---

**Human Subjects or Proprietary Information**

Does your project involve the potential use of human subjects? (Y/N): N

Does your project involve the potential use of proprietary company information? (Y/N): N

---

**STUDENT SIGNATURE**

**By signing and submitting this form, you acknowledge** that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

---

**TO BE FILLED BY A COURSE INSTRUCTOR**

**The capstone topic is approved by a course instructor.**



**COURSE INSTRUCTOR'S NAME AND SIGNATURE:**

---

**COURSE INSTRUCTOR APPROVAL DATE:**

**Project Compliance with IRB (Y/N):**    Y

