Write up in Microsoft Word, Apple Pages, LaTeX, or alternative word processor with fixed pages.

The content of each section may be in a different order that listed in the rubric, but all items must be included. If you are concerned about the language portion, consider utilizing our university's resources such as the Writer's Workshop.

Format [20 Points]

Content	Full Points		No Points	
≤ 1 page w/ reasonable font size and margins	10	yes	0	no
Section headers for: Overview, Data, and Expectations	5	all three	0	two or fewer
Typos/grammatical errors	5	none to some	0	many

Unless in a table or list, please use complete sentences. PLP

Overview [20 Points]

In a paragraph...

- Describe your **prediction problem**: What is the label you want to predict?
- Provide **context**: Help an intelligent person who only knows "ECON 101" understand your project.
- Provide **motivation**: Why should an economist or policy-maker care?

Content	Full Points		Half Points		No Points	
Prediction problem	10	correct	-15	X & y relationship	-40	causal question
Context	5	provided			0	missing
Motivation	5	convinced	2.5	not convinced	0	missing

Data [40 Points]

- Is your prediction problem a **regression or classification problem**?
- From where is your data (provide a URL if data are from competition websites such as Kaggle)?
- What is the structure of your data: **cross-section, panel, or time series**?

  Observations

  - What is your **unit of observation** (*count not necessary*)?
  - What is your **time dimension**? From when to when?
  - What is your **total number of observations** and is it **more than 2,000**?[1]

  Features

  - Do you have **eight or more** features?[2]
  - Do **features match label** observation and time unit?
  - In a table or list over features (1) what is the feature, (2) is it continuous or categorical, (3) what is its unit of observation,[3] and (4) what is its time dimension?

[1] You may have fewer if you have prior approval from instructor.
[2] You only need to show the first eight if the page limit is an issue.
[3] If you have **time series** data, you can simply write TS for unit of observation of your features. You may also have up to two lags of your label as features.

Hints:

	Cross-section	Panel	Time Series
Unit of Observation	30,000 individuals	3,142 regions	one financial asset
Time Dimension	December 1913	annual: 2010–2019	daily: 6/31/2007 to 9/15/2008
n	30,000	31,420	413 (*this is not enough*)

Because my label (weight in lbs) is continuous, I am studying a regression problem. My data are exclusively from IPUMS MEPS. I have a total of 40,000 individuals from 2010–2012. My data has the strucutre of a _____. My features are:

1.	height of the individual	continuous	individual	annual
2.	their age	continuous	individual	annual
⋮	⋮	⋮	⋮	⋮
10.	race	categorical	individual	annual
11.	female	categorical	individual	annual

Content	Full Points		No Points	
Regression or classification	5	correct	0	incorrect
Data source(s)	2.5	provided	0	missing
Data structure	2.5	correct	0	incorrect
Observations				
Unit of observation	2.5	correct	0	incorrect
Time dimension (what & when)	2.5	correct	0	incorrect
n	5	$\geq$ 2,000	0	$<$ 2,000
Features				
8 features w/ 4 descriptions (name, continuous/categorical, unit of obs., time dimension)	20	0.5 points per valid feature-description pair		

Expectations [20 Points]

Content	Full Points		No Points	
In a table or list similar to features...				
Top 3 strongest predictors	6	2 points each		
Their correlation with your label	6	2 points each		
In a sentence...				
What is one potential problem (of any kind) you may encounter?	8	provided	0	missing

Hint:

1.	height	postive
2.	age	concave quadratic
3.	female	negative