

# CAPSTONE PROJECT APPENDIX

(DATA DICTIONARY)



ANA BEATRIZ POTJE  
APPLIED DATA SCIENCE BOOTCAMP  
CAMBRIDGE SPARK

SUPERVISED BY RICARDO MONTI / KIRE KOLAROSKI  
LONDON, 26<sup>TH</sup> MARCH 2020

## Appendix A – Data Dictionary

Feature Name	Description	DataType
<b>TARGET_deathRate</b>	<b>Dependent variable. Mean per capita (100,000) cancer mortalities(a)</b>	<b>float64</b>
avgAnnCount	Mean number of reported cases of cancer diagnosed annually(a)	float64
avgDeathsPerYear	Mean number of reported mortalities due to cancer(a)	int64
incidenceRate	Mean per capita (100,000) cancer diagnoses(a)	float64
medIncome	Median income per county (b)	int64
popEst2015	Population of county (b)	int64
povertyPercent	Percent of populace in poverty (b)	float64
studyPerCap	Per capita number of cancer-related clinical trials per county (a)	float64
binnedInc	Median income per capita binned by decile (b)	object
MedianAge	Median age of county residents (b)	float64
MedianAgeMale	Median age of male county residents (b)	float64
MedianAgeFemale	Median age of female county residents (b)	float64
Geography	County name (b)	object
AvgHouseholdSize	Mean household size of county (b)	float64
PercentMarried	% of county residents who are married (b)	float64
PctNoHS18_24	% of county residents ages 18-24 highest education attained less than high school (b)	float64
PctHS18_24	% of county residents ages 18-24 highest education attained high school diploma (b)	float64
PctSomeCol18_24	% of county residents ages 18-24 highest education attained some college (b)	float64
PctBachDeg18_24	% of county residents ages 18-24 highest education attained bachelor's degree (b)	float64
PctHS25_Over	% of county residents ages 25 and over highest education attained high school diploma (b)	float64
PctBachDeg25_Over	% of county residents ages 25 and over highest education attained bachelor's degree (b)	float64
PctEmployed16_Over	% of county residents ages 16 and over employed (b)	float64
PctUnemployed16_Over	% of county residents ages 16 and over unemployed (b)	float64
PctPrivateCoverage	% of county residents with private health coverage (b)	float64
PctPrivateCoverageAlone	% of county residents with private health coverage alone (no public assistance) (b)	float64
PctEmpPrivCoverage	% of county residents with employee-provided private health coverage (b)	float64
PctPublicCoverage	% of county residents with government-provided health coverage (b)	float64
PctPublicCoverageAlone	% of county residents with government-provided health coverage alone (b)	float64
PctWhite	% of county residents who identify as White (b)	float64
PctBlack	% of county residents who identify as Black (b)	float64
PctAsian	% of county residents who identify as Asian (b)	float64
PctOtherRace	% of county residents who identify in a category which is not White, Black, or Asian (b)	float64
PctMarriedHouseholds	% of married households (b)	float64
BirthRate	Number of live births relative to number of women in county (b)	float64

(a) Years 2010-2016

(b) 2013 Census Estimates