Tackling Teenage Pregnancy

Sandy Lee & Laura Suchomska

Flatiron School

Problem Framing

- In 2017, 194,377 babies were born to women aged 15–19 years.
- A Record low for USA teens, and a drop of 7% from 2016.

The Importance of Prevention

- Pregnancy and birth contribute to school dropout rates among girls.
- Only 50% of teen mothers finish high school
- The children of teenage mothers are more likely to be low achievers
  - 1991 - 2015 saw a 64% drop in teen births, which resulted in $4.4 billion in public savings in 2015 alone.

(source: Centers for Disease Control and Prevention, 2020)

Dataset

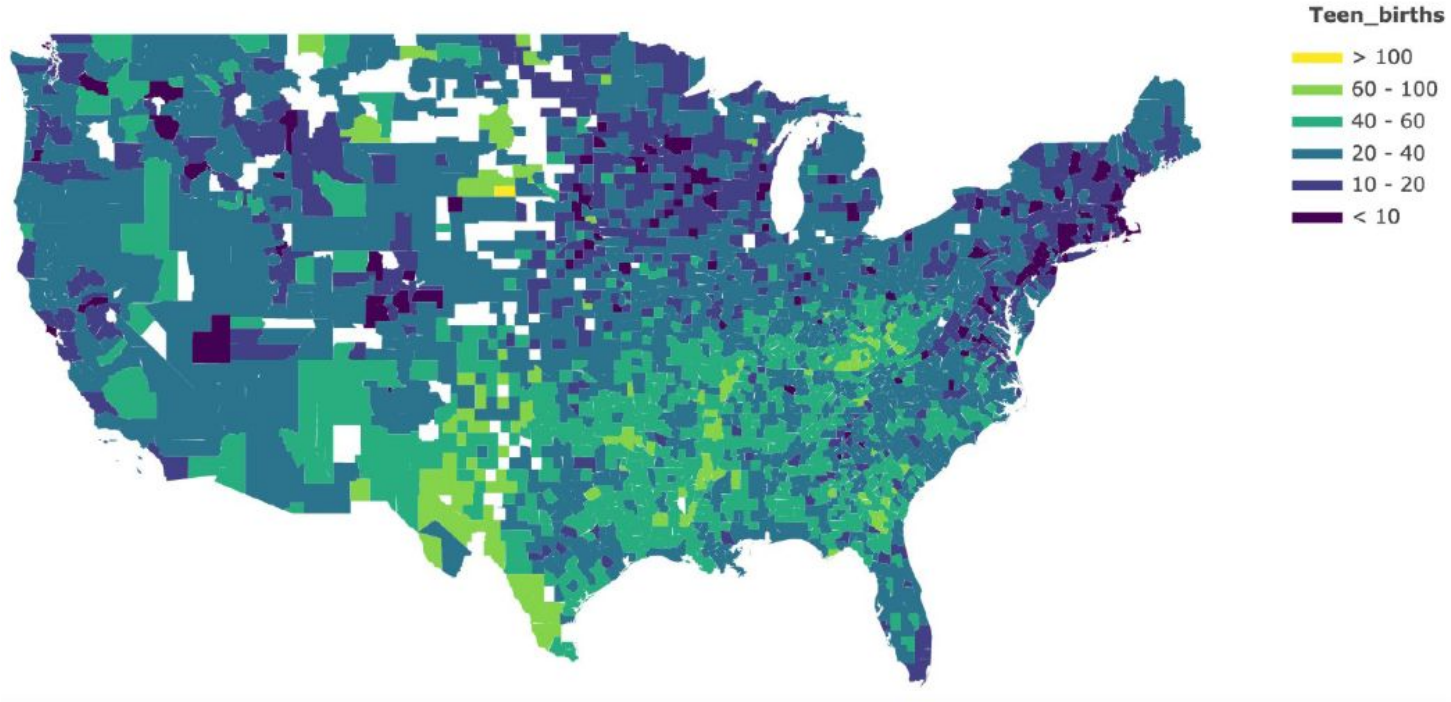
- The Data was gathered in a collaboration between the University of Wisconsin and the Robert Wood Johnson Foundation.
- The rankings and the data information used in this project were obtained from the 2019 County Health Ranking.
- The dataset comprises 3142 rows, 25 columns out of 534.

Stakeholders

Health and Education Departments

Mapping Teenage pregnancy in the US

Map showing the teen birth rates according to their geographical incidence



County Health Ranking 2019

Teenage women giving birth (15-19 years of age)

Counties with the highest rates	State	Counties with the lowest rates	State
Todd	SD	Hunterdon	NJ
Kusilvak Census Area	AK	Hampshire	MA
Brooks	TX	Whitman	WA
Benson	ND	Tolland	CT
Roosevelt	MT	Norfolk	MA

What machine learning is and isn’t

Pattern Recognition

- 40 different models
- 3 different algorithms
- 325 different factors examined
- Top 10 contributing factors identified

Where to spend your budget

Health & Food Security	10% impact
University Education/Rural Environments	8% impact
Income Inequality for over 18s	7% impact
Low birthweight & income inequality	6% impact
Access to exercise/Rural Environments	>5% impact

*estimated

Recommendations

Education is key. We strongly advise the health and education departments to invest their resources on a series of educational interventions, including:

- Increasing the quality of mainstream education
- Sexual education
- Focus on educating teenagers regarding cooking and eating behaviour
- Increase the access to physical activities in rural areas

If you have been affected by the topics raised in this presentation, please access the following links for more information and help:

NHS - Teen pregnancy help

https://www.nhs.uk/conditions/pregnancy-and-baby/teenager-pregnant/

PLANNED PARENTHOOD

https://www.plannedparenthood.org/learn/teens/preventing-pregnancy-stds/i-think-im-pregnant-now-what

Thank you !

Appendix 1 - Teen Birth rates per county

	Teen_births	State_Abbreviation	Name
1816	2.394971	NJ	Hunterdon County
1248	2.522548	MA	Hampshire County
3040	3.312009	WA	Whitman County
323	3.365701	CT	Tolland County
1251	3.558332	MA	Norfolk County
3144	3.627570	WI	Ozaukee County
1820	3.724538	NJ	Morris County
2298	3.730379	PA	Centre County
301	3.895184	CO	Pitkin County
1902	4.058654	NY	Putnam County
891	4.151100	IA	Story County
2838	4.375000	UT	Morgan County
1808	4.391089	NJ	Bergen County
2353	4.517515	RI	Bristol County
270	4.700110	CO	Douglas County

	Teen_births	State_Abbreviation	Name
2465	110.236220	SD	Todd County
85	96.399100	AK	Kusilvak Census Area
2592	95.549738	TX	Brooks County
2029	94.070696	ND	Benson County
1669	93.750000	MT	Roosevelt County
2456	88.397790	SD	Oglala Lakota County
2413	87.971275	SD	Buffalo County
1628	85.144388	MT	Big Horn County
1845	84.841441	NM	Luna County
2763	83.006769	TX	Reeves County
2425	82.486551	SD	Dewey County
2627	81.889764	TX	Deaf Smith County
2440	80.967403	SD	Jackson County
1499	80.188679	MS	Tunica County
2791	79.956585	TX	Terry County
178	79.574161	AR	St. Francis County

Appendix 2 - Data Preparation

- Run Multicollinearity tests to identify most strongly correlated features and removed them.

- Manually removed columns, based on their clear irrelevance to the objectives of the projects, example: Air pollution, flu vaccinations, etc.

- Missing values replaced by the features' median.

```
In [32]: 1 to_drop

Out[32]: ['Poor physical health days raw value',
          'Poor mental health days raw value',
          'Adult smoking raw value',
          'Children in poverty raw value',
          'Percentage of households with high housing costs',
          'Life expectancy raw value',
          'Premature age-adjusted mortality raw value',
          'Frequent physical distress raw value',
          'Frequent mental distress raw value',
          'Diabetes prevalence raw value',
          'Food insecurity raw value',
          'Limited access to healthy foods raw value',
          'Uninsured adults raw value',
          'Uninsured children raw value',
          'Median household income raw value',
          'Children eligible for free or reduced price lunch raw value',
          'Severe housing cost burden raw value']
```

Appendix 3 - Baseline Model

In [94]:
```
baseline_model = cross_validate(regression, X_train, y_train, scoring='r2', cv=crossvalidation, return_train_score = True)
baseline_model
```

Out[94]:
```
{'fit_time': array([0.00598454, 0.00299239, 0.00299191, 0.00199485, 0.00299191,
        0.00199437, 0.00199461]),
 'score_time': array([0.0009954 , 0.0009973 , 0.00099754, 0.00099802, 0.00099707,
        0.00099635, 0.00099826]),
 'test_score': array([0.71069167, 0.68081211, 0.66257228, 0.68086107, 0.62099895,
        0.63297934, 0.69230138]),
 'train_score': array([0.67299491, 0.67792233, 0.68185384, 0.67825863, 0.68689286,
        0.68464373, 0.67636968])}
```

Dep. Variable:	Teen births raw value	**R-squared:**	0.670
Model:	OLS	**Adj. R-squared:**	0.668
Method:	Least Squares	**F-statistic:**	317.4
Date:	Wed, 26 Feb 2020	**Prob (F-statistic):**	0.00
Time:	09:24:13	**Log-Likelihood:**	-11177.
No. Observations:	3142	**AIC:**	2.240e+04
Df Residuals:	3121	**BIC:**	2.252e+04
Df Model:	20		
Covariance Type:	nonrobust		

Appendix 4 - Interactions and Polynomials

Top 7 interactions: [('Poor or fair health raw value', 'Food environment index raw value', 0.681), ('Low birthweight raw valu
e', 'Children in single-parent households raw value', 0.681), ('Food environment index raw value', 'Children in single-parent h
ouseholds raw value', 0.681)]

```
degree 2 {'fit_time': array([0.05791736, 0.01994467, 0.01994443, 0.02194118, 0.02393818,
       0.01794982, 0.02094579]), 'score_time': array([0.00199628, 0.00099754, 0.00099754, 0.00099754, 0.        ,
       0.0009973 , 0.        ]), 'test_score': array([0.54523904, 0.29460553, 0.70333745, 0.69403275, 0.62460998,
       0.61987627, 0.7285961 ]), 'train_score': array([0.80371706, 0.80179695, 0.80188278, 0.79979824, 0.80318526,
       0.79986126, 0.79940538])}
degree 3 {'fit_time': array([2.91320992, 2.95410013, 2.84040594, 2.81846166, 2.71474051,
       2.72172141, 2.73369002]), 'score_time': array([0.00299215, 0.00199533, 0.00299311, 0.00199509, 0.00199533,
       0.00199819, 0.00199676]), 'test_score': array([-147.13690624, -116.46414553,   -9.42994679,  -83.48363155,
       -396.9758179 , -104.59810616,  -28.14929125]), 'train_score': array([0.98037315, 0.97878919, 0.97800496, 0.97960806, 0.9
7904812,
       0.97887337, 0.97766655])}
degree 4 {'fit_time': array([6.31511211, 6.15254593, 5.81943989, 5.49131346, 5.70075607,
       6.35899448, 6.27422094]), 'score_time': array([0.01096964, 0.00997496, 0.00897598, 0.00798345, 0.00897503,
       0.00797915, 0.01097059]), 'test_score': array([-6.95526345e+02, -3.35296979e+04, -1.50118253e+01, -3.00764140e+02,
       -2.91038123e+04, -2.44909781e+03, -7.18226504e+01]), 'train_score': array([1., 1., 1., 1., 1., 1., 1.])}
```

Contributing Features

Graph showing the features contributing to higher incidences of teen births