DubsTech Datathon 2026

Website: https://access-to-care.streamlit.app/

Access to Care Over Time: Modeling Delayed Medical Care by Age Group

Team: Lily
Members: Leelavathi Raja

Motivation

Topic: Delayed getting medical care due to cost among adults"

Subgroup: Start with "18-34 years", later included "50-64 years", "65 years and older".

Outcome: % of adults delaying medical care due to cost.

Time period: 2019–2024 (annual)

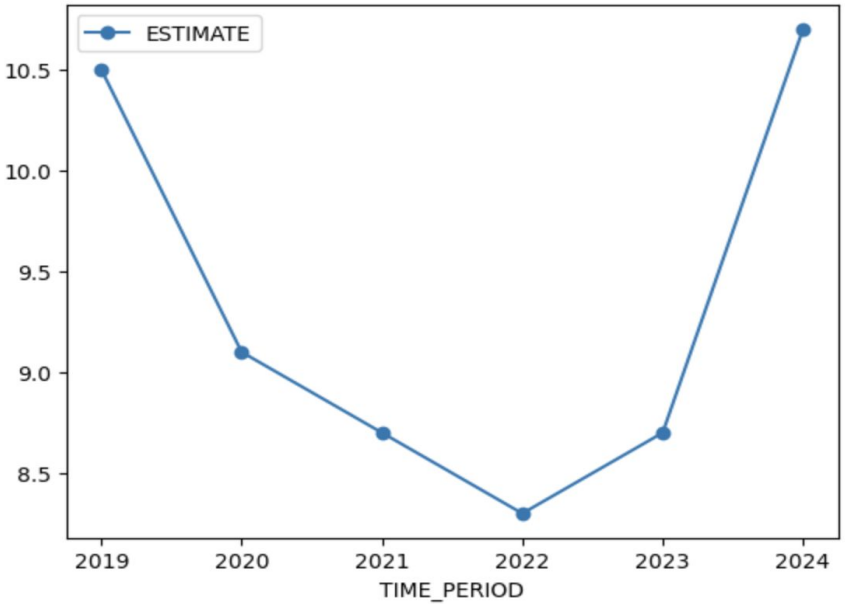
Subgroups analyzed: 18–34 years 50–64 years 65 years and older

Preprocessing - preprocessing.ipynb

- **Explored the Columns**
- **Retained only relevant variables:** 'TOPIC', 'GROUP', 'SUBGROUP', 'TIME_PERIOD', 'ESTIMATE', 'FLAG'
- Retained only Reliable Rows:
  - Removed missing or invalid entries using the ‘FLAG’ column
  - This FLAG indicated the reliability of the ESTIMATE column.
  - This FLAG is missing then ESTIMATE is reliable.
  - If FLAG contains data, then ESTIMATE is unreliable.
- Ensured TIME_PERIOD is numeric for time-based modeling.

Exploratory analysis - model.ipynb

- **Filtered to a single health topic:** Delayed getting medical care due to cost among adults
- Selected age-based subgroups: 18–34 years



- Decline during early pandemic years

- Rebound in later years

- Nonlinear patterns varying by age group

Modeling - model.ipynb

1. **Linear trend:** Assumes constant rate of change

   Estimate = intercept + coefficient*t

2. **Piecewise linear:** trend Allows slope change at 2022.

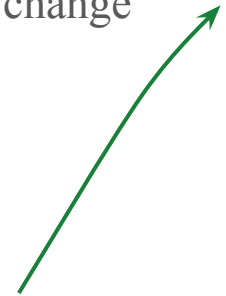
   Estimate = intercept + coefficient1*t + coefficient2* post_2022

   Here, post_2022 = 1 if t >= 2022 , 0 else

3. **Quadratic trend:** Captures smooth decline and rebound

   Estimate = intercept + coefficient1*t + Coefficient2*(t)^

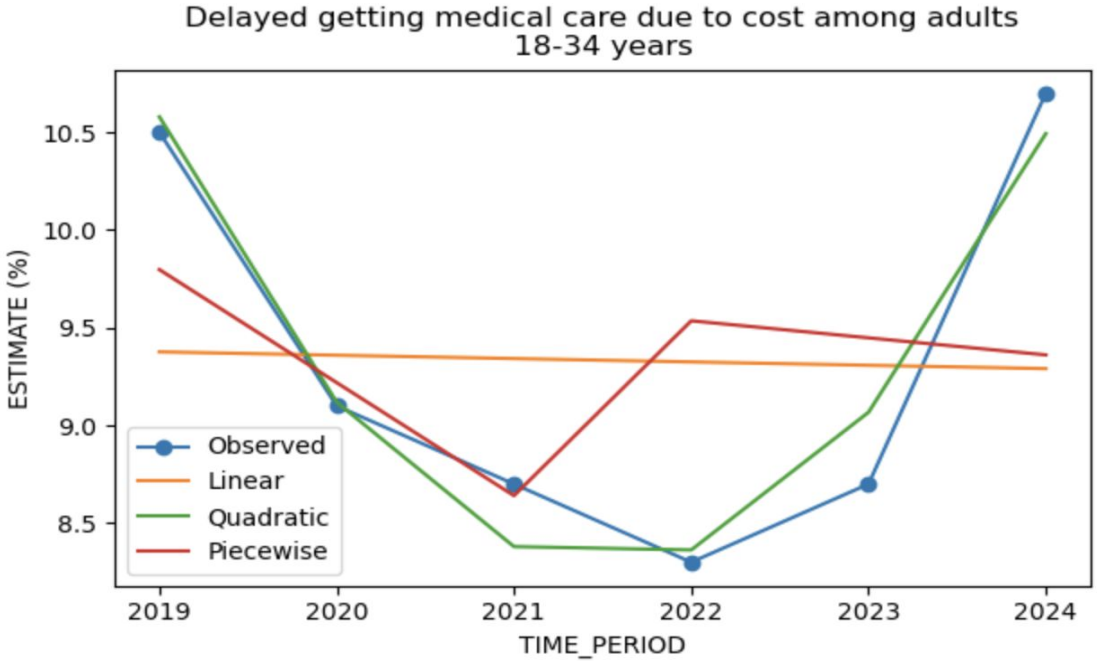
- Improves numerical stability
- Simplifies interpretation:
Intercept = baseline year estimate
Coefficients = annual change



Note: Converted calendar year to centered time index: t = TIME_PERIOD − min (TIME_PERIOD)

Evaluation Metrics

How models were compared using R^2, MSE, MAE (Metrics computed with training data)



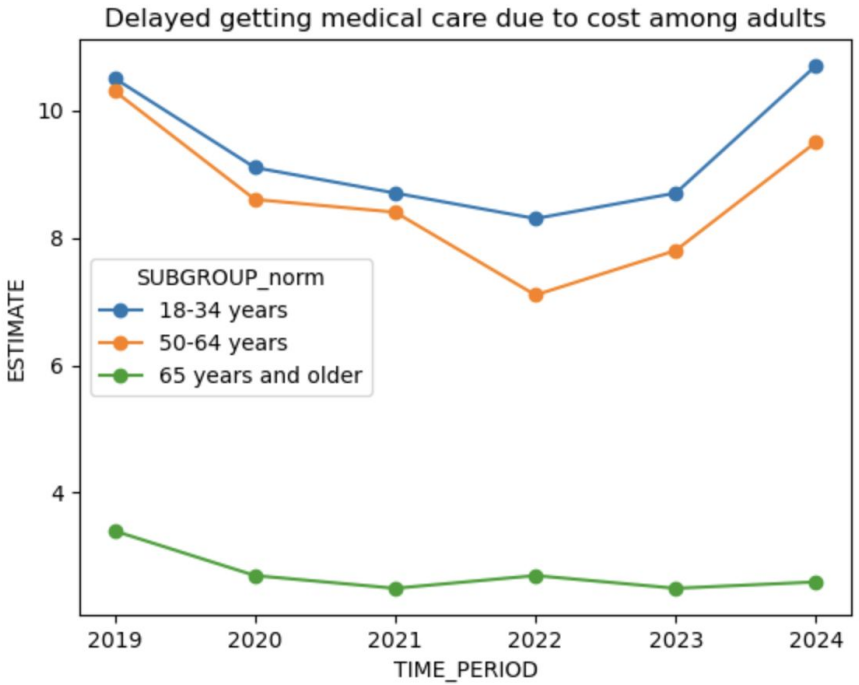
Evaluation Results

```
'R2_linear': 0.00099796710404764 79,
'R2_piecewise': 0.1479713042455 6015,
'R2_quadratic': 0.9436148586213271,
'mse_lin': 0.8580317460317454,
'mse_piecewise': 0.7317979797979798,
'mse_quadratic': 0.04842857142857129,
'mae_lin': 0.8444444444444444,
'mae_piecewise': 0.7006060606060608,
'mae_quadratic': 0.17571428571428557,
```

“Quadratic model”
explained more variance, have smaller MSE(0.048) & MAE(0.1757)

For 2025: 12.64% Delayed care for the selected population, reflecting unmet healthcare needs.

Subgroups Comparisons:



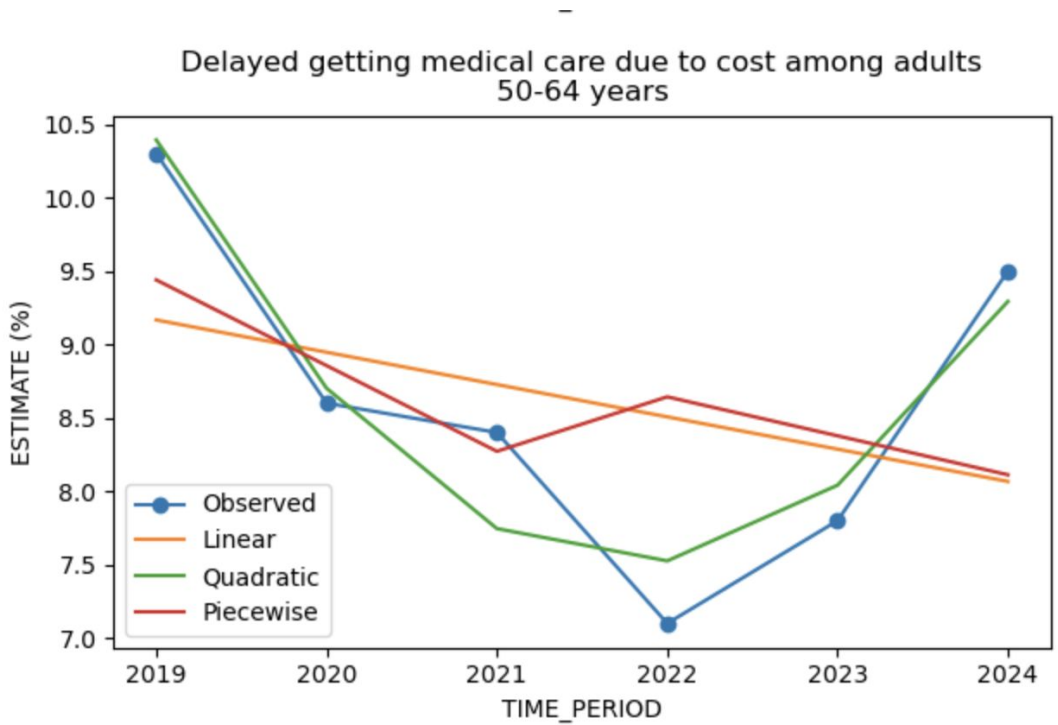
- All age groups show a decline in delayed care during 2019–2022, with the largest drop among ages 50–64

- A rebound is observed after 2022, particularly for ages 18–34 and 50–64

- Adults aged 65+ consistently report substantially lower levels of delayed care.

- The relative ordering of age groups remains stable over time.

Model Comparisons: 50-64 years

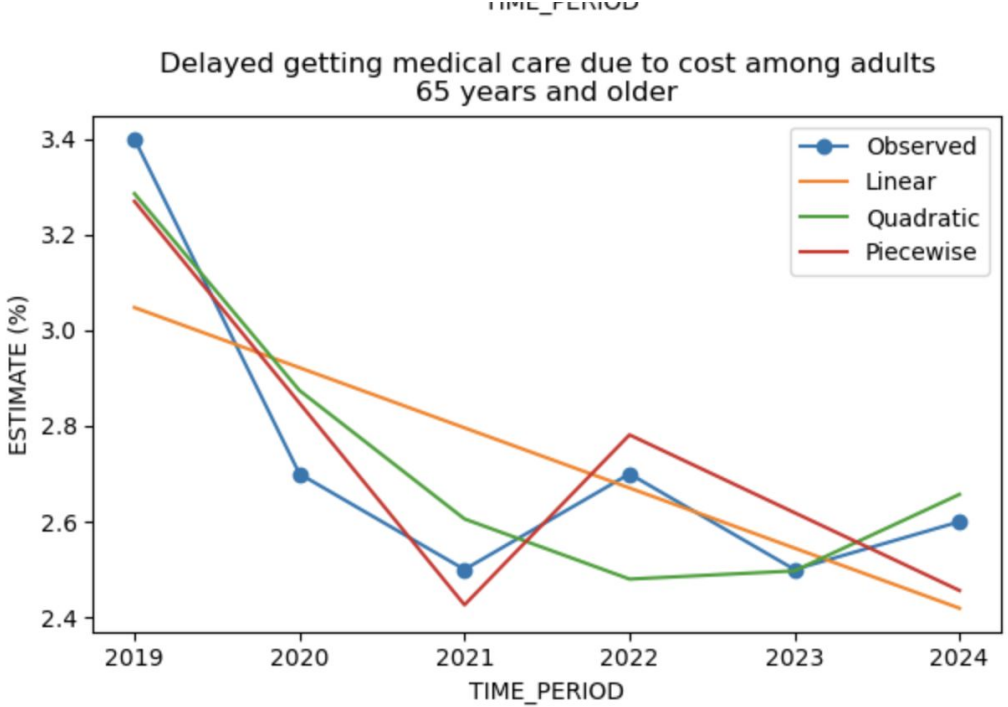



```
'R2_linear': 0.12778476238370606,
'R2_piecewise': 0.17578805403799125,
'R2_quadratic': 0.8899529437120588,
'mse_lin': 0.9635555555555565,
'mse_piecewise': 0.910525252525253,
'mse_quadratic': 0.12157142857142839,
'mae_lin': 0.8555555555555556,
'mae_piecewise': 0.7921212121212124,
'mae_quadratic': 0.2871428571428565,
```

"Quadratic model"
explained more variance,
have smaller MSE(0.1215)
& MAE(0.2871)

For 2025: 11.28% Delayed care for the selected population, reflecting unmet healthcare needs.

Model Comparisons: 65 years and older




```
'R2_linear': 0.4823920265780731,
'R2_piecewise': 0.850369978858351,
'R2_quadratic': 0.8146179401993356,
'mse_lin': 0.04946031746031745,
'mse_piecewise': 0.014297979797979787,
'mse_quadratic': 0.017714285714285703,
'mae_lin': 0.18761904761904757,
'mae_piecewise': 0.11606060606060598,
'mae_quadratic': 0.11238095238095232,
```

"Piecewise model"
explained more variance, have smaller MSE(0.014)

For 2025: 2.29% Delayed care for the selected population, reflecting unmet healthcare needs.

Conclusion

- Linear trends are inadequate.
- Nonlinear models substantially improve fit.

Limitation:

- There were only 6 points for the chosen TOPIC & SUBGROUP.