

# *Data-Driven Decision Making* in Nursing Home Investments

BANA 620 | Data Mining & Predictive Analytics for Business |  
DJK Predictive Partners: Denise Becerra | Josh Dennis | Kanak Sharma  
May 7, 2024







# 01 *Introduction*

## Nursing Facilities in The U.S

As the Baby Boomer Generation ages, reputable nursing home facilities with dedicated teams of healthcare professionals has become a necessity in the United States to support the 76 million who are in need of rehabilitation, hospice, or round the clock care service.

### **In 2020 alone:**

- Number of Nursing Homes: 15,300 (CDC)
  - Number of licensed beds: 1.6 million (CDC)
  - Number of residents: 1.3 million (CDC)
  - Proportion of Nursing Homes with for-profit ownership: 70.3% (CDC)
- 
- 

## 02 Methodology

- Through data collection from Centers for Medicare & Medicaid Services and the use of data analysis & predictive modeling techniques, such as Multi-Linear Regression, Lasso Regression, and K - Nearest Neighbor, **DJK Predictive Partners** proposes the analytical objective:

*What factors impact the profitability of nursing homes pre & post COVID-19 pandemic?*

- Answering this objective can aid in strategic decision making in the healthcare industry like potential resource allocations, performance monitoring, and help gain a competitive positioning advantage over others in the healthcare industry.



## 03 Data Description

### Data Shape:

- 28 Features
- 102,420 observations

### Target Variable:

- *Net Income*

### Independent Variables:

- Key
- Year
- Provider\_CCN
- Number\_of\_Beds
- SNF\_Admissions\_Total
- Total\_Discharges\_Total
- Gross\_Revenue
- Inpatient\_PPS\_Amount
- Inpatient\_Revenue
- Less\_Total\_Operating\_Expense
- Net\_Income
- Net\_Income\_from\_patients
- Net\_Patient\_Revenue
- Overhead\_Non\_Salary\_Costs
- Salaries\_wages\_and\_fees\_payable
- Total\_Costs
- Total\_General\_Inpatient\_Revenue
- Total\_Income
- Total\_Salaries\_From\_Worksheet\_A
- Total\_Salaries\_adjusted
- Wage\_related\_Costs\_core
- Total\_Assets
- Number Of Certified Beds
- Provider Type
- Overall Rating
- Total Weighted Health Survey Score
- Total Amount Of Fines In Dollars
- Covid\_Period

### Percent of Missing Values:

- 3.62% from original dataset
  - (Removed Total 3,849 NA's)

## 04

# Data Preprocessing Findings

- **Facilities in Rural vs Urban**
  - Rural: 72.8% of facilities
  - Urban: 27.2% of facilities
- **Control Types**
  - Proprietary Corporation: ~85%
  - Proprietary-Other: ~5%
  - Voluntary Nonprofit-Other: ~4%
  - Proprietary-Partnership: ~3%
  - Voluntary Nonprofit-Church: ~2%
- **Highest # of Beds by States**
  - Florida (Large Population of Retirees)
  - Texas
  - California
  - New York
  - Ohio
  - Illinois
- **Precovid vs Post Covid**
  - >Pre-Covid Net Income Avg: Significantly Higher
  - >Post-Covid Net Income Avg: Significantly lower
- **Medicare vs Medicaid Accepting Facilities:**
  - Medicare & Medicaid: ~ 60,000 locations
  - Medicaid only: < 10,000 locations
- **Top Fine Paying States**
  - Texas
  - California
  - Ohio
  - Florida
  - Illinois



## 05 *Modeling Results*

Multi-Linear  
Regression

### Model Accuracy

#### Training

R-Squared: 0.933263  
MSE: 6.356 B  
MAE: 6.872 B

#### Testing

R-Squared: 0.924417  
MSE: 6.960 B  
MAE: 6.840 B

Lasso  
Regression

### Model Accuracy

#### Training

R-Squared: 0.933260  
MSE: 6.357 B  
MAE: 6.862 B

#### Testing

R-Squared: 0.924430  
MSE: 6.959 B  
MAE: 6.826 B



K-Nearest  
Neighbor




### Model Accuracy

#### Training

R-Squared: 0.8940  
MSE: 10.096 B  
MAE: 19.136 B

#### Testing

R-Squared: 0.873211  
MSE: 11.676 B  
MAE: 20.582 B







## 06 RESULTS

- DKJ acknowledges potential limitations of the model in predicting a continuous variable due high bias in data and non-linear relationship among prediction and predictor variables.
- Income subsets and overall expenses were identified as influential factors.
- Future research may benefit from narrowing the sample scope to include only data closely related to the property type of interest for prospective buyers.
- On comparing model performance, we realized that best model to work for this problem could be :
  - Random Forest Regression: Handles nonlinear relationships well between features. Robust to outliers and less affected by multicollinearity.
  - Neural Network Regression: Can capture complex relationships in the data and useful for handling nonlinearities.



# 07

## Recommendations

Based on Analysis & Findings

1. Payroll-related expenses demonstrated the most significant influence on Net Income compared to other expense categories.
2. Insight suggests potential investment opportunities in areas with lower average salaries for healthcare professionals.
3. Senior housing facilities can be a fruitful investment, suggesting the potential for generating positive net income
4. Some properties observed in the analysis are largely successful, implying they generate healthy net income
5. External forces, such as a pandemic, can have disastrous impacts on the financial standing (including net income) of a property



## 08 *References*

- <https://www.cdc.gov/nchs/fastats/nursing-home-care.htm>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1464018/>
- <https://www.britannica.com/topic/baby-boomers>
- [https://www.cdc.gov/nchs/pressroom/nchs\\_press\\_releases/2024/20240525.htm#:~:text=The%20general%20fertility%20rate%20in,consistently%20decreased%20by%202%25%20annually.](https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2024/20240525.htm#:~:text=The%20general%20fertility%20rate%20in,consistently%20decreased%20by%202%25%20annually.)
- <https://www.epa.gov/smartgrowth/smart-growth-and-affordable-housing>
- <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-detail.html>

# Thank You

Denise Marie Becerra | Josh Dennis | Kanak Sharma