DATA ANALYTICS FOR HEALTHCARE

NAME – SWAPNIL NICOLSON DADEL

https://github.com/swapitsneil/healthcareanalysis

Project Overview

The project examined the MetroHealth83 dataset (originally 83 rows, filtered to 77 after cleaning) to evaluate healthcare safety hazards in 77 cities in the U.S. The research utilized Power BI as the workhorse tool, with selected tasks as a part of Tableau, to benefit from 17 columns such as RateMDs, RateBeds, PctChangeMedicare, and engineered features (newHAI, newPDI, newMedicareDependencyRatio). The aim was to discover disparities, forecast events, and offer actionable suggestions for public health improvement.

Phase 1: Data Preparation and Acquisition

Task 1: Load the Health Dataset with Appropriate Attributes

Imported MetroHealth83.csv into Power BI, verifying 16 columns (City, NumMDs, RateMDs, etc.) and 83 rows.

Task 2: Assess Dataset's Completeness, Quality, and Relevance

Complete with no missing values, of high quality with no errors, and relevant for healthcare analysis. Problems involved 6 duplicate cities and no Population column.

Task 3: Clean Initial Data

Deleted 6 duplicate cities (77 rows remaining), capped RateMDs outliers at 510 (99th percentile) and RateBeds at 550, approximated Population as a Whole Number, and standardized formatting. Saved as MetroHealth83_Cleaned.pbix.

Phase 2: Data Exploration and Visualization

Task 4: Explore the Dataset Using Power BI

Generated readable visuals in Power BI: Histogram for RateMDs (binned at 50, Y-axis < 20 cities, easy title/labels) demonstrating distribution of physician rates, table for important metrics (City, NumMDs, etc.), matrix for summary statistics (min/max/average). Investigated structure and distributions in 77 cities.

Task 5: Visualize Data Patterns and Relationships

Graphicized patterns in Power BI: Bar chart ("Number of Physicians per 100,000 - Lowest Rates by City") identifies low RateMDs (<150, e.g., Hanford-Corcoran, CA), line chart ("Impact of Physician Availability - Medicare Change by City") indicates negative PctChangeMedicare (e.g., -2.2% in Scranton), enhanced scatter plot ("Investment Opportunities: Physician and Bed Availability by City") plots low RateMDs (<200) and RateBeds (<300) for business consideration. Identified prospective safety risks from restricted healthcare access.

Task 6: Detect and Visualize Possible Outliers and Anomalies

Detected outliers in Power BI: RateBeds > 585 (e.g., Shreveport, 590) visualized using scatter plot and table with conditional formatting, truncated at 585. No new

RateMDs anomalies (truncated at 510). (Note: Altered from 550 to 585 based on 99th percentile for accuracy.)

Phase 3: Data Analysis and Feature Engineering

Task 7: Perform Univariate and Bivariate Analyses

Performed univariate analysis in Power BI: Histogram ("Univariate: Physician Rate Distribution") indicates RateMDs distribution (peak 150–300), card ("Avg Hospital Beds per 100,000") shows RateBeds average (~300). Bivariate analysis: Scatter plot ("Bivariate: Physician vs. Bed Rates") indicates correlation, matrix ("Bivariate: Physician Rates vs. Medicare Change") correlates low RateMDs with negative PctChangeMedicare, indicating safety incident risks. (Adapted to Tableau with similar visuals.)

Task 8: Create New Features Using Feature Engineering

Added new measures in Power BI: newHAI (average of RateMDs and RateBeds, limited to 400), newPDI (per capita of physicians, limited to 1000), and newMedicareDependencyRatio (Medicare % of population, limited to 15). Visualized as bar chart ("Healthcare Access Index by City"), line chart ("Population Density Impact on Physicians"), and scatter plot ("Medicare Dependency vs. Physician Rates"), which improved incident prediction and risk analysis. (Recreated in Tableau using calculated fields.)

Task 9: Implement Outlier Detection Techniques

Implemented outlier detection in Power BI: Flagged RateBeds > 550 (capped at 550), newHAI > 400, newPDI > 1000, and newMedicareDependencyRatio > 15 as outliers through DAX calculated columns within Table View. Capped new features at corresponding limits for data accuracy. (Modified to Tableau with calculated field caps.)

Phase 4: Reporting and Documentation

Task 13: Summarize Findings and Insights

Summarized conclusions: Data Acquisition and Preparation (Tasks 1–3) loaded and cleaned data, Data Exploration (Tasks 4–6) identified RateMDs distribution and safety hazards, Data Analysis and Feature Engineering (Tasks 7–9) validated correlations and created new features. Insights of significance are cities with low newHAI (<200) and high newPDI (>500) are exposed to safety risks, and high newMedicareDependencyRatio (>15%) with low RateMDs showing at-risk groups.

Task 14: Design Impactful Data Visualizations

Developed compelling data visualizations in Power BI through refining current dashboards: Revised histogram ("Safety Impact: Physician Rate Distribution Across Cities") emphasizes low-rate risks, combo chart ("Critical Safety Risks: Low Physician Rates vs. Medicare Decline") indicates correlated decreases, scatter plot ("Strategic Dashboard: Healthcare Access vs. Population Density") identifies investment opportunities, and matrix ("Safety Correlation: Physician Rates vs.

Medicare Trends") connects safety trends, presenting intricate safety patterns in a clear way. (Reconfigured to Tableau dashboards.)

Task 15: Prepare an In-Depth Report

Documented analysis: Phase 1 collected and sanitized data, Phase 2 discovered patterns, Phase 3 analyzed and created features (newHAI = (RateMDs + RateBeds)/2, newPDI = Population/NumMDs, newMedicareDependencyRatio = (NumMedicare/Population)*100) through DAX, truncated outliers. Salient findings: Low RateMDs and RateBeds are associated with risks; newHAI < 200 and newPDI > 1000 point to under-resourced cities. Suggestions: Invest in low newHAI cities, neutralize adverse PctChangeMedicare regions, track newMedicareDependencyRatio, reverify caps every year.

Phase 5: Real-World Applications

Task 16: Apply Insights and Visualizations

Applied Power BI insights and visualizations to solve actual health issues: Histogram reveals cities such as Hanford-Corcoran, CA with poor RateMDs (<150), indicating healthcare shortages. The combo chart points out Scranton with dismal PctChangeMedicare (-2.2%), reflecting dwindling access, triggering special interventions. The scatter plot employs newHAI and newPDI to target investment in resource-short cities (e.g., low newHAI <200, high newPDI >500), filling rural healthcare gaps. (Modified to Tableau with maps.)

Task 17: Measure the Effect of Data-Driven Recommendations

Measured the effect of data-driven recommendations on safety operations: Investment recommendations in low newHAI (<200) and high newPDI (>500) cities improve resource distribution, minimizing safety risks. Reducing cities with negative PctChangeMedicare and low RateMDs (e.g., Scranton) by doctor recruitment enhances response to emergencies. Tracking newMedicareDependencyRatio (>15%) aids in proactive care, making informed decisions possible by offering actionable priorities.

Project Conclusion and Future Steps

Task 18: Project Conclusion

Ended the project with major findings: Low RateMDs (<150) and RateBeds (<300) associate with safety hazards, being augmented by adverse PctChangeMedicare, whereas newHAI <200 and newPDI >500 indicate under-resourced communities, and elevated newMedicareDependencyRatio (>15%) indicates vulnerability. Importance lies in presenting a data-driven approach to healthcare disparities management, resource optimization, and safety improvement, with a scalable solution for public health planning.

Task 19: Discuss Potential Future Extensions

Covered possible future expansions: Incorporating other data sources such as hospital wait times, patient outcomes, or socioeconomic data might improve risk assessments. Adding real-time data (e.g., through API feeds) would allow for dynamic tracking of newHAI and newPDI. Developing machine learning models to forecast safety events given newMedicareDependencyRatio trends would facilitate better decision-making. Improvements involve setting outlier cut-offs (e.g., 510, 550) with greater data sets and testing Population estimates.

Overall Project Achievements

Tools Utilized: Mainly Power BI for cleansing, analysis, and visualization of the data

Deliverables: Cleaned data, exploratory charts, predictive attributes, effective dashboards, and actionable advice.

Impact: Offered a solid framework to detect and avert healthcare safety hazards with possible real-world implementation and enlargement.