ARCHDIOCESE OF SEATTLE

MSDS 25.2

Archdiocese of Seattle

Planning and Mission Effectiveness Demographic Data History

6 June 2025

Sponsor Organization – Archdiocese of Seattle

Liaisons – Elijah Morgan (Director of Planning & Mission Effectiveness) &

Jo Singleton (Data Analyst)

Faculty Advisor – Dr. Ariana Mendible

Team Members

Janani Krishnamurthy

Lakshmi Prasanna Kumar Nalabothu

Satwik Kunaparaju

Ting-Yu Lin

Suryakailash Ramesh

Executive Summary

This project aims to leverage data-driven insights to help the Archdiocese of Seattle increase engagement within parish communities. By analyzing demographic, financial, and sacramental data, the project aims to identify key trends and factors influencing the growth of cultural communities and young adult participation. Predictive modeling, data visualization, and spatial analysis techniques uncover insights into parish dynamics, engagement patterns, and community growth. These insights will guide the development of strategic plans, optimize resource allocation, and strengthen community outreach. The project will provide actionable recommendations that ensure the archdiocese's ministries effectively meet the needs of a diverse population, foster greater engagement, and support long-term sustainability, benefiting parish leaders and community members. Based on past parish population trends, future participation has been estimated for efficient pastoral planning in the future. Parish participation is influenced by racial and marital diversity, as well as the presence of specific cultural groups. Higher engagement is seen in areas with a mix of African American and White Non-Hispanic populations and varied marital status among Mass attendees. Additionally, parishes with significant Caribbean, Southeast Asian, and Middle Eastern populations show distinct participation patterns, reflecting cultural segmentation. Growth trends, particularly the involvement of young adults, have also been analyzed. Furthermore, participation across different language groups has been compared to assessing how it has increased or decreased over time. The analysis reveals that while overall parish registrations decline, parishes offering Masses in languages like Spanish and Vietnamese show stronger growth based on the T-Test. Similarly, parishes transitioning from mission to parish status experienced stronger outcomes in these areas, including higher sacramental activity, greater household registration, and increased participation among young adults and cultural communities. This showed that when people were able to worship in a language, they felt comfortable with, they became more connected and involved in parish life. Analysis found something important, parishes that offered Mass in more than one language, especially Spanish and Vietnamese, had much higher levels of participation. Young adult involvement is rising, but some outreach opportunities remain untapped during population surges. Structural changes, such as a mission to parish transitions, are linked to higher engagement, while parish-to-mission shifts show fewer impacts. Parishes with non-English Masses consistently report higher participation, though financial support is similar. These insights help the archdiocese plan more inclusive and effective strategies tailored to the needs of diverse communities.

TABLE OF CONTENT

1. Introduction

Catholicism is one of the largest religions in the world and has an important position in the United States. However, Catholic affiliation has seen a slight decline over time. According to the Pew Research Center, 20% of U.S. adults will identify as Catholic in 2024. This percentage has remained stable since 2014 but is slightly down from 24% in 2007 [1]. In Washington State, the proportion of Catholics is even lower, with recent estimates indicating that 14% of the state's adult population identifies as Catholic [2].

The Archdiocese of Seattle (AOS) is a geographic jurisdiction of the Catholic Church that has 174 locations including 136 parishes, missions, and stations. It has 74 Catholic schools offering faith-based education. It serves over one million Catholics in Western Washington and is responsible for providing religious services, education, and support for various cultural communities [3]. The Archdiocese of Seattle launched Partners in the Gospel to revitalize Catholic engagement in 2024. This long-term pastoral planning effort aims to strengthen faith communities and ensure parish sustainability.

This initiative places particular emphasis on two key demographic groups: cultural communities and young adults. Western Washington's Catholic population is diverse and is made up of cultural communities that share a common ethnic, cultural, or linguistic background. These communities enrich parish life with traditions, multilingual Masses, sacraments, and cultural events. Meanwhile, young adults, defined as individuals aged 18 to 39, are critical to the Church's future and form a key part of the parish community. Understanding the factors that influence their participation and developing effective pastoral strategies will be key to strengthening their connection with the Church.

This project aims to analyze Catholic population trends, evaluate socioeconomic factors affecting parish dynamics, and provide data-driven insights to support strategic pastoral planning within the Archdiocese of Seattle. Understanding these trends is vital for planning parish mergers, allocating clergy, and sustaining community programs. The analysis primarily covers the period from 2001 to 2024 and draws on internal records on parish demographics, household counts, sacramental participation, and parish finances. U.S. Census [4] data was also incorporated to provide a broader view of the age and community composition of parish neighborhoods. While the Pew Research Center has examined national religious trends using scenario-based demographic projections [5], such approaches rarely apply advanced analytics at the local level. This study seeks to bridge that gap by applying forecasting, clustering, and spatial analysis to better inform pastoral planning in the Archdiocese of Seattle. Forecasting population trends may help predict parish viability over the next decade, while clustering and spatial tools can identify shared needs across neighboring parishes and communities.

2. Problem Statement

The Archdiocese of Seattle seeks to gain a deeper understanding of the historical and current trends in the growth, participation, and engagement of two key demographic groups: cultural communities and young adults. These insights will help shape future pastoral planning efforts, ensuring that the AOS remains vibrant, sustainable, and responsive to the needs of its diverse population.

Firstly, this study focuses on cultural communities. Among communities, defined by shared ethnic, cultural, or linguistic backgrounds, the archdiocese explores how different groups have grown over time and how that growth influences parish life in areas such as sacramental participation, financial contributions, and involvement in ministry and events. The study also considers how decisions such as parish closures, site relocations, ministry changes, or the introduction of Masses in new languages have helped or made things harder for these communities. It also studies which cultural groups have shown strong growth. By identifying key growth trends and assessing the responsiveness of cultural ministries to demographic changes over time, the analysis seeks to provide insights into the future composition and needs of parish populations.

Similarly, Young adults are an important group for the future of the AOS. Their level of involvement can vary, and factors like participation in sacraments and attendance at mass play a role in this. By studying how young adults are engaging with parishes in the Archdiocese of Seattle, the analysis will help identify ways to improve outreach and build a stronger connection between young adults and their faith communities.

By conducting this analysis, the Archdiocese of Seattle aims to better understand the factors driving parish growth, participation, and decline. The goal is to strengthen community engagement, improve outreach especially to cultural communities and young adults and ensure that ministries and programs are aligned with the current and future needs of the Catholic population. This study will also support smarter decision-making when it comes to allocating resources and responding to demographic changes across the region. The insights gained will guide future pastoral strategies and help the Church foster a renewed sense of belonging, connection, and active participation within all parish communities.

3. Methodology

This section outlines the methodology for analyzing parish population trends, financial stability, and sacramental participation using a combination of machine learning models, statistical methods, and visualization techniques. By applying predictive modeling, clustering, and dimensionality reduction techniques, the approach captures key patterns, identifies influential factors, and enables data-driven decision-making for parish planning.

3.1 Data Sources

The data for this study was sourced from Excel files provided by the Archdiocese of Seattle. These files included household counts, parish demographic details, sacramental records from 2000 to 2024, and financial data from 2018 to 2024. The demographic data captured individual-level information such as age, gender, marital status, ethnicity, preferred language, membership type, start date, and parish affiliation. Sacramental records included counts of baptisms, first communions, confirmations, marriages, and funerals. Financial records provided detailed information on ordinary income, revenue, and expenses at the parish level. Together, these datasets offered a comprehensive view of parish activity over time, enabling analysis of trends in membership, participation, and financial health across different communities.

To provide additional population context and support demographic comparisons, we also incorporated U.S. Census data on young adults in Western Washington from 2010 to 2022.

3.2 Data Preparation

These datasets were structured and stored in SQLite[6] by creating different tables to enable efficient querying and analysis. An additional table was created to aggregate overall data, including average and total counts across all years, alongside demographic details. This table facilitates analyses focused on broader trends without year-based variations. Each row in the dataset represents a specific parish each year, capturing key metrics such as household counts, sacraments performed (baptisms, confirmations, marriages, funerals, and communions), mass attendance, and financial contributions. Additionally, parish-level demographic insights are included. A Parish ID is a unique identifier assigned to each parish. Some parishes lacked an official Parish ID, so identifiers within the 900 series were generated to maintain uniformity.

During preprocessing, missing values were handled systematically. If a parish did not exist before a certain year, the missing values in its records were replaced with zeroes. For other missing entries, values were imputed using the average of the previous and next year's data to preserve trend consistency. Upper age limits were set at 110, and rows with invalid age values were removed. Membership records, which refer to individuals officially registered under a parish as Catholics along with their registration dates, were reviewed. Any records with future dates were flagged as errors and adjusted to align with the parish's establishment year. For statistical modeling techniques such as Principal Component Analysis (PCA) and K-means clustering, the data was scaled using the average mass attendance per parish. This value served as a realistic estimate of the active parishioner base and ensured that variations across parishes were normalized before applying the models.

To enhance analysis, demographic data from the U.S. Census was incorporated to evaluate the presence and participation of young adults in parish communities. The census data was selected based on geographic relevance and included the following population metrics:

- Total Population
- Age-Specific Groups: Males and females aged 18-19, 20-24, 25-29, 30-34, and 35-39
- Spanish and Asian, Pacific Islander [7]

This dataset was integrated for further modeling and visualization, helping to assess market share trends and parish engagement over time

3.3 Modeling Techniques

Multiple models are applied to analyze parish-level impacts and population trends over time. Each model was designed to address a specific analytical goal.

3.3.1 Time Series Forecasting (ARIMA)

The AOS wants to estimate the future parish population Trends. A time-series model developed using ARIMA [8], a method particularly suited for stationary data [9], where statistical properties such as mean, variance, and autocorrelation [10] remain constant over time and no seasonality is present.

The registered parish household data, although not initially stationary, was transformed through differencing, and the data does not have any seasonality.

The ARIMA model has three key parameters [11] that are tuned:

- p (autoregressive): the number of lag observations included in the model.
- d (differencing): the number of times the data is different to make it stationary.
- q (moving average): the number of lagged forecast errors that are included in the prediction equation.

The model had been evaluated by checking if the forecasted values follow a logical trend based on past behavior. For example, if historical data shows a consistent decline but the model predicts a sudden rise, it may indicate a flaw in the forecasting logic. The accuracy of the forecasts is assessed using the percentage difference between the actual values and the forecasted values, focusing on the last two data points.

The dataset includes nearly 25 years of parish participation data, from 2001 to 2024. Each entry spans between two calendar periods (e.g., 2000–2001), but for simplicity, the data is labeled using the ending year (e.g., labeled as 2001). The model uses data from 2000 to 2022 for training, reserving the last two years as test data.

This model helps forecast future parish population trends and serves as a foundation for strategic pastoral planning. However, it does not account for external influences such as declining religious engagement or temporary demographic shifts, which could affect the accuracy of long-term projections.

3.3.2 Principal Component Analysis (PCA) for Dimensionality Reduction

PCA [12] is an unsupervised learning technique used for dimensionality reduction, particularly effective in high-dimensional datasets with multicollinearity. It was employed to identify key

demographic and financial variables influencing changes in parish populations over time. By transforming correlated variables into a smaller set of uncorrelated principal components, PCA simplifies the dataset while preserving the most critical variance-based information. All numeric variables were standardized using Scikit-learn's "Standard Scaler" to ensure equal contribution across features. For PCA, the variables included primarily consisted of normalized or per-capita metrics, such as counts per mass attendance (e.g., young adult count per mass, male/female/adult/children count per mass), and various averages per mass attendance including sacraments like baptisms, communions, confirmations, funerals, marriages, and household participation. Financial metrics, such as average income, revenue, and expenses per mass attendance, were also excluded due to their potential to disproportionately influence the variance structure, thereby overshadowing demographic patterns. Based on the explained variance ratio, the top 6 components were retained, achieving a cumulative variance threshold of at least 60%, which was considered sufficient for reliable insights. The top 6 features contributing to each principal component were extracted using absolute loading values, highlighting the dominant factors driving variation in parish demographic and financial patterns. To uncover latent groupings among parishes, K-Means [13] clustering was applied to PCA-transformed data. This allowed efficient segmentation of parishes based on their compressed representations. The number of clusters was determined using domain knowledge and visualization tools such as the elbow method. PCA effectiveness was evaluated through explained variance and interpretability of loadings rather than accuracy metrics.

3.3.3 Lasso

To understand young adult participation across different deaneries, Lasso regression [14] was applied to identify the most relevant predictors. As a regularized linear model using an L1 penalty, Lasso automatically selects features by shrinking the coefficients of less important or redundant variables to zero. This method is particularly useful when dealing with multicollinearity and large number of input variables. A 5-fold cross-validation procedure was used to tune the regularization parameter and ensure model generalizability. Variables with non-zero coefficients were retained for further analysis. While Lasso simplifies the model and highlights representative features, it may exclude variables that are relevant but highly correlated with others.

3.3.4 Ordinary Least Squares (OLS)

OLS regression [15] is an optimization approach used to find the best-fitting straight line in a linear regression model. It helps examine how each variable is related to young adult participation. The model estimates the relationship between each feature and the target variable (registered young adults / total registered parishioners) while controlling for the effects of other variables. It produces p-values, which help identify which features are statistically significant. The limitation of OLS is that it does not include regularization, so it is sensitive to multicollinearity when features are highly correlated.

3.3.5 T-Test

The Independent Samples T-test [16] is a statistical method used to determine whether there is a significant difference in the means of two independent groups. It helps assess whether an intervention or attribute (like language Mass offerings or parish status change) is associated with

measurable outcomes (such as attendance or sacraments). We performed right-tailed t-tests to evaluate whether Group A outperforms Group B on key metrics, focusing specifically on whether values in Group A are significantly greater.

The data used for the T-test analysis was sourced from the overall Parish data table within the SQLite database. In preparing the dataset for statistical analysis, all non-numeric fields and identifying columns were excluded to ensure that only measurable and comparable variables were evaluated. The analysis focused specifically on numeric indicators such as attendance ratios, sacramental participation, and financial performance metrics. Since the T-test is concerned with comparing group means and not sensitive to feature scale, scaling or normalization of variables was not required for this statistical test.

To analyze the impact of language-specific Mass offerings, two groups were defined. Group A included parishes that offered Masses in languages other than English, while Group B included parishes that offered only English-language Masses. Furthermore, to explore the effects of parish status changes, three groups were created. Group A consisted of parishes that had been converted from parish to mission, Group B consisted of those that transitioned from mission to parish, and Group C included all other parishes that had not undergone any conversion. The comparisons were conducted between Group A and Group C, and between Group B and Group C, to identify whether changes in parish status had measurable impacts.

To evaluate statistical significance, a right-tailed independent samples T-test was performed for each numeric feature. This approach tested whether values in Group A or B were significantly greater than those in Group C (or the comparison group).

A threshold of $p \leq 0.05$ was used to determine statistical significance. After running the tests, the results were categorized. Metrics that were significantly higher in the experimental group (e.g., Group A) compared to the control group (e.g., Group B or C) were flagged as significant. Metrics with p-values above the threshold were marked as non-significant and retained for completeness.

3.4 Tools

A variety of tools were used throughout this project to support data analysis and visualization. Python, primarily through Google Colab, was used for data cleaning, transformation, and statistical modeling. SQLite was used to structure and store datasets efficiently for querying. ArcGIS supports spatial analysis and mapping of parish-level trends.

4. Results

4.1 Trends in Household Registration Rates Over Time and Future Projections

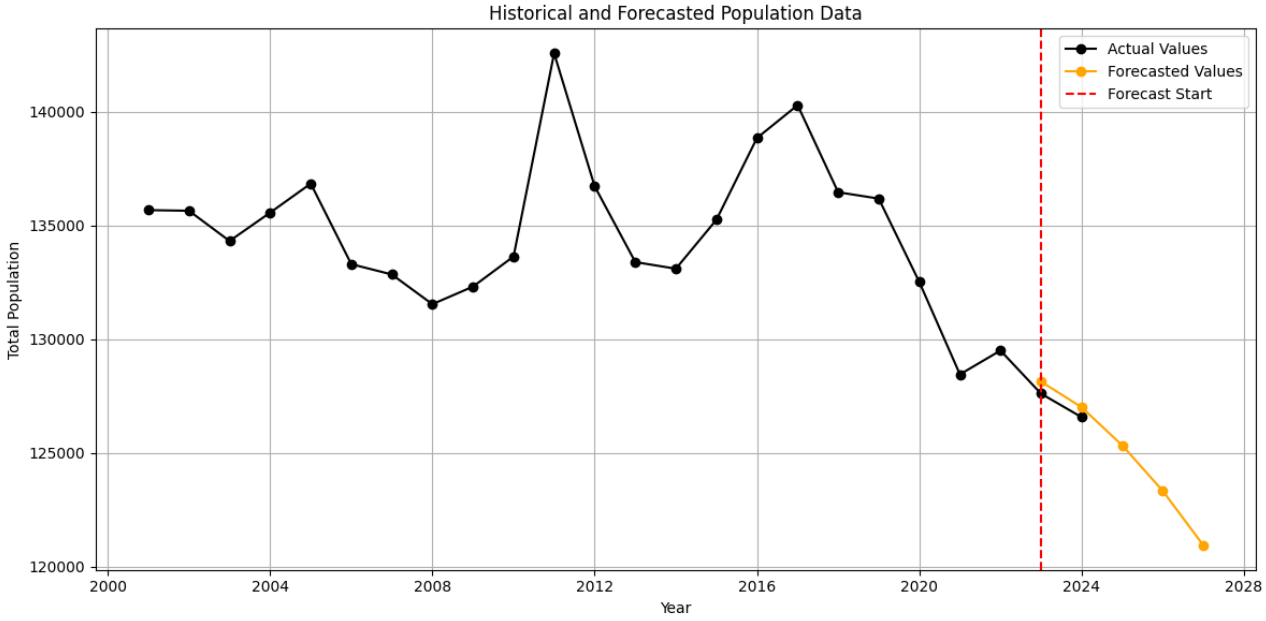


Figure 1. Trends and Forecasts of Registered Households in Western Washington

The above graph illustrates the past, present, and projected future Parish population in Western Washington from 2000 to 2027. The data from 2000 to 2022 is used as training to estimate the next 5 years' data with the values in the years 2022 and 2023 used for validation. The overall parish population has been declining over the past two decades. The projected population continues this downward trend, with approximately 1,877 Household expected to decrease each year until 2027. In the next 3 years i.e. till 2027, an estimate of 5,633 Household is projected to decrease with a 4.45% drop overall.

4.2 Unique Growth Trends in Cultural Communities.

As shown in Figure 1 above, the overall trend of Registered parish count is declining.

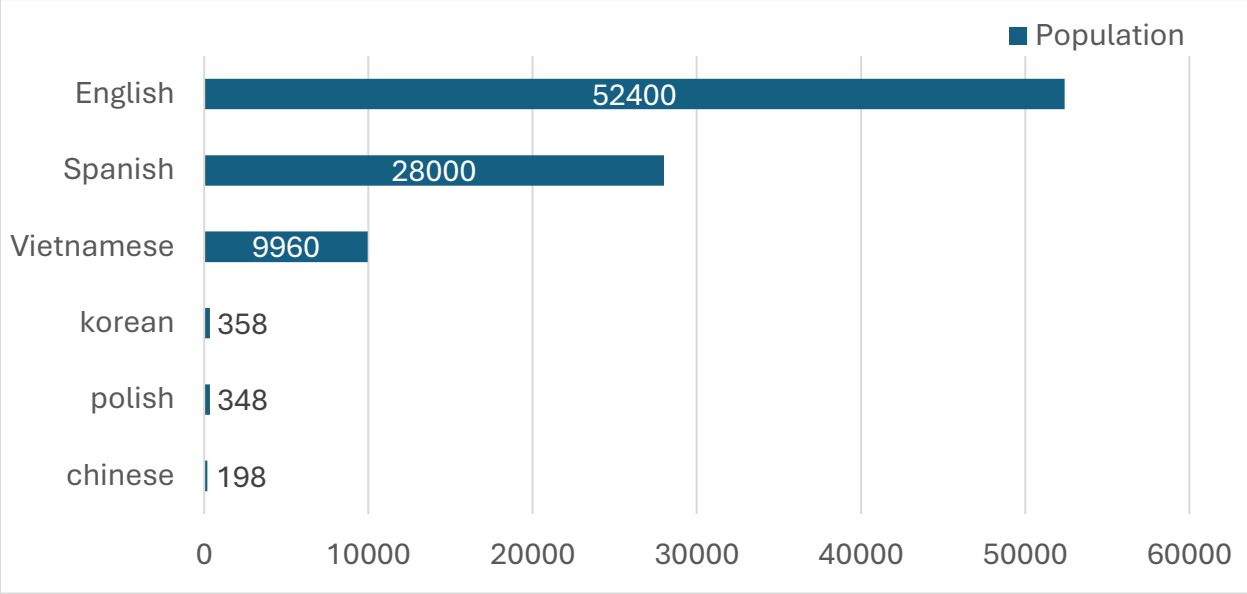


Figure 2. Language-wise Population Distribution in the Parish

The chart illustrates the population distribution in a parish based on the spoken language. English is the dominant language, spoken by 52,400 individuals, followed by Spanish with 28,000 speakers. Vietnamese is the third most common language, with a population of 9,960. In contrast, the number of people speaking Korean, Polish, and Chinese is significantly smaller, with 358, 348, and 198 individuals respectively. This indicates that the parish is primarily English-speaking, with notable Spanish and Vietnamese communities, while other language groups represent a much smaller portion of the population.

However, the Spanish and Vietnamese masses-offering parishes have shown a noticeable amount of growth over this time frame. Figures 3 and 4 illustrate the Vietnamese and Spanish registration trends in a similar time frame.

Figure 3. Registration (Population) Trends in Parishes Offering Spanish Mass

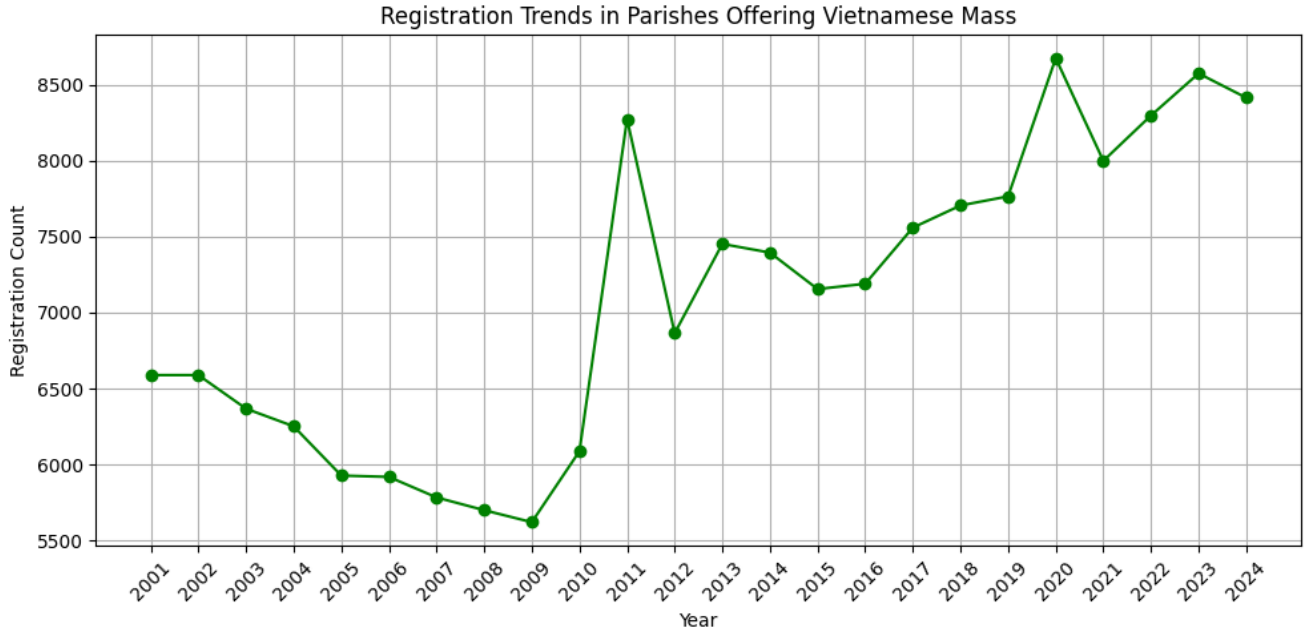


Figure 4. Registration (Population) Trends in Parishes Offering Vietnamese Mass

Registrations in the parishes that offer Vietnamese Masses have seen a growth of 27.7% from 2001 to 2024. Whereas Spanish Mass offering parishes have seen a growth of 7.48% growth in a similar time span. Contrary to this observation the total registered household for the entire organization has experienced a decrease of 6.71%. However, if Vietnamese and Spanish mass offering parishes are removed from the overall trend, The registration count decreased further by approximately 10.69% from 2001 to 2024, the trend is shown below in figure 5.

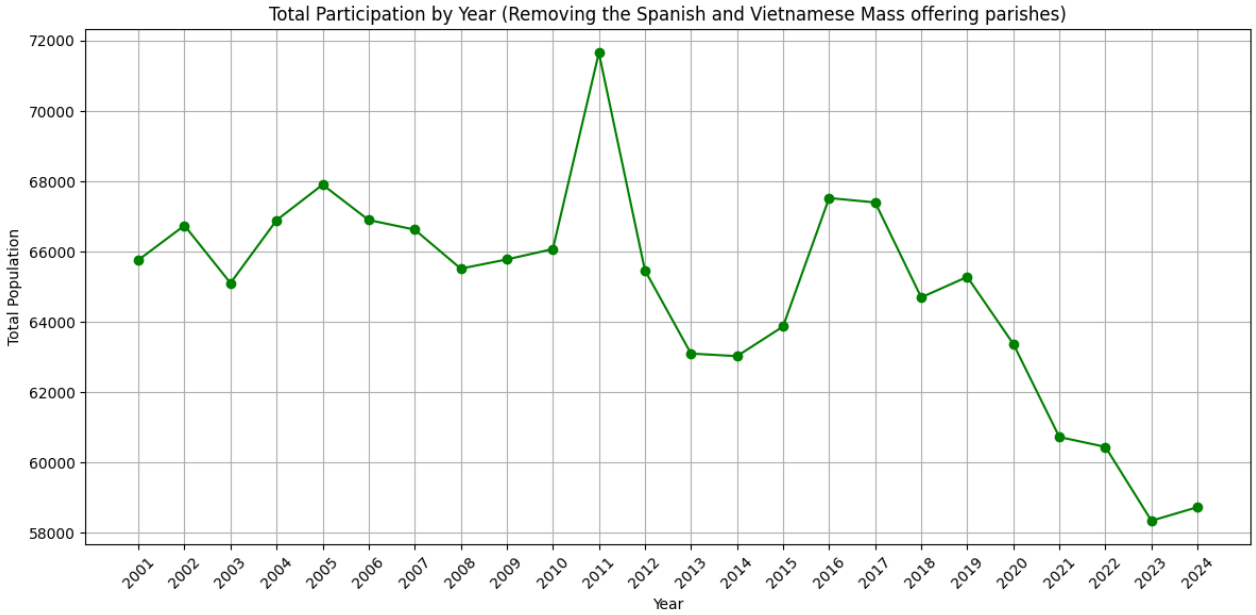


Figure 5. Total Registrations (Population) Excluding Spanish and Vietnamese Mass–Offering Parishes

To compare all the three graphs, Figure 6 was plotted as shown below, where all the values are scaled between 0 and 1, They are scaled relative to their own individual minimum and maximum values. Rolling average is applied to the data with a window of over 5 years.

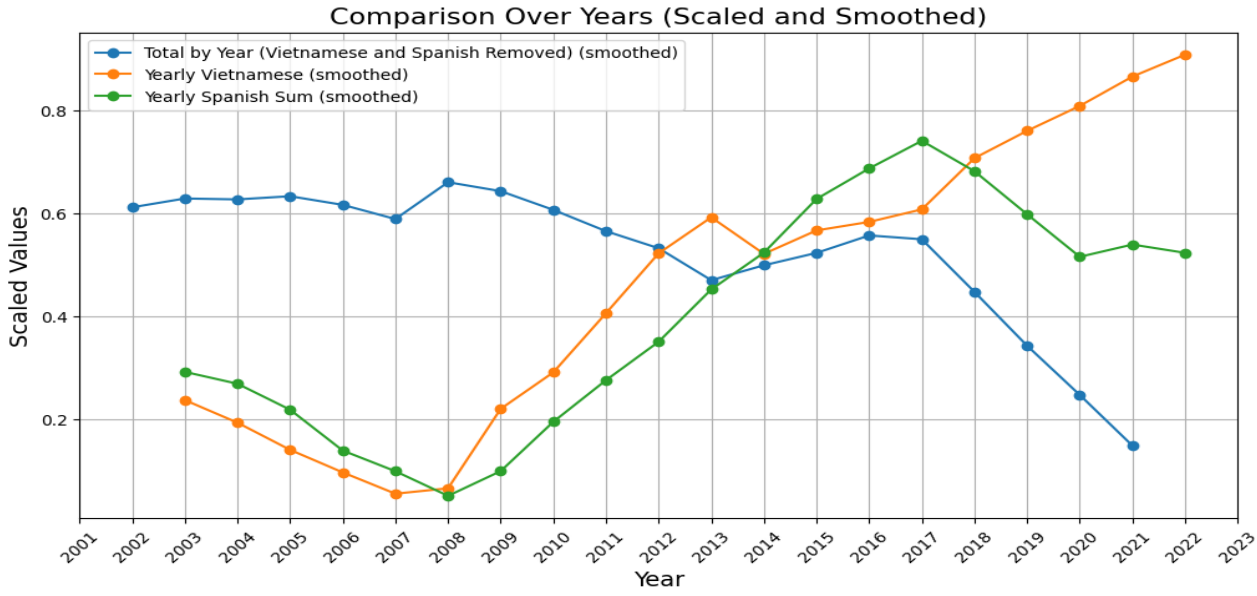


Figure 6. Comparison of Trends Among Vietnamese, Spanish, and the Total Population (Excluding Spanish and Vietnamese Speakers)

4.3 Analyzing Parish Participation Rate Among Young Adults

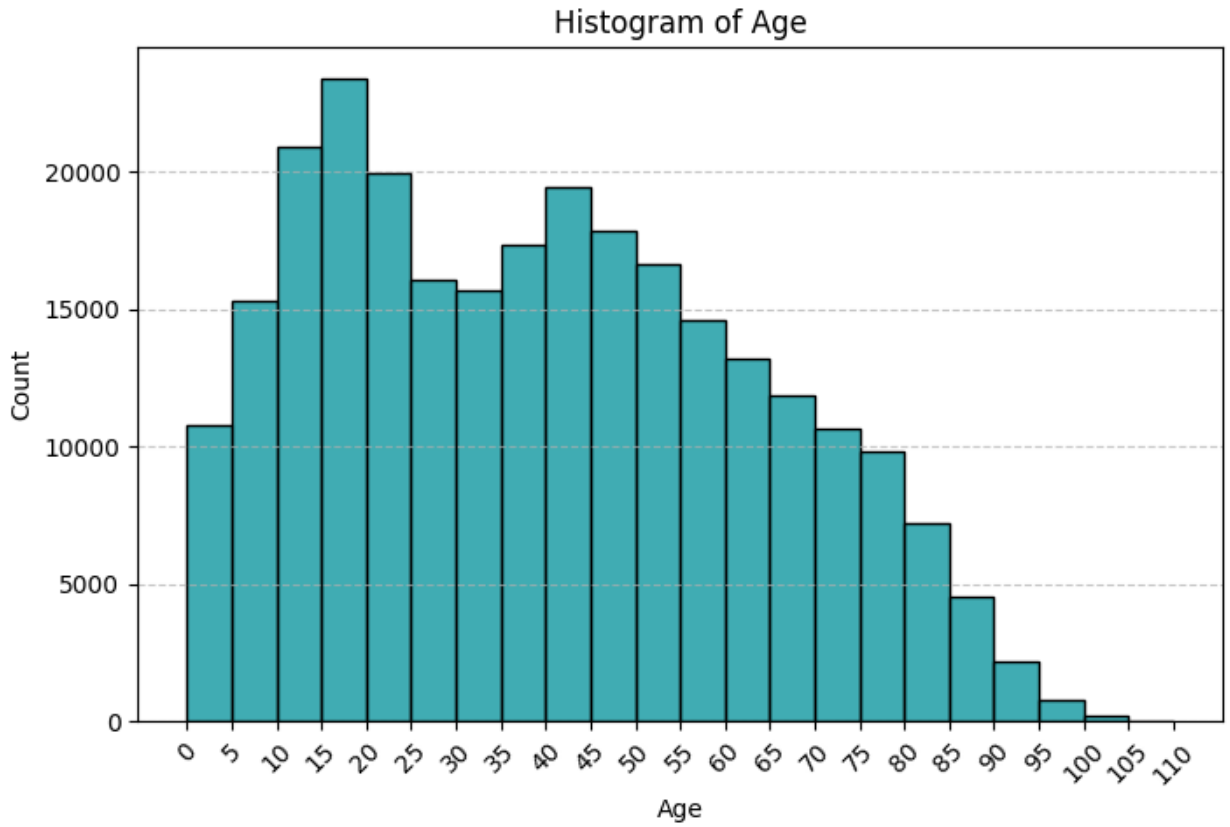


Figure 7. Age Distribution of Registered Parishioners

This histogram displays the age distribution of registered parishioners, ranging from 0 to 110 years old. It highlights the population structure within the parish community. Notably, young adults aged 18 to 39 account for 29% of all registered parishioners, representing a substantial and vital segment of the parish.

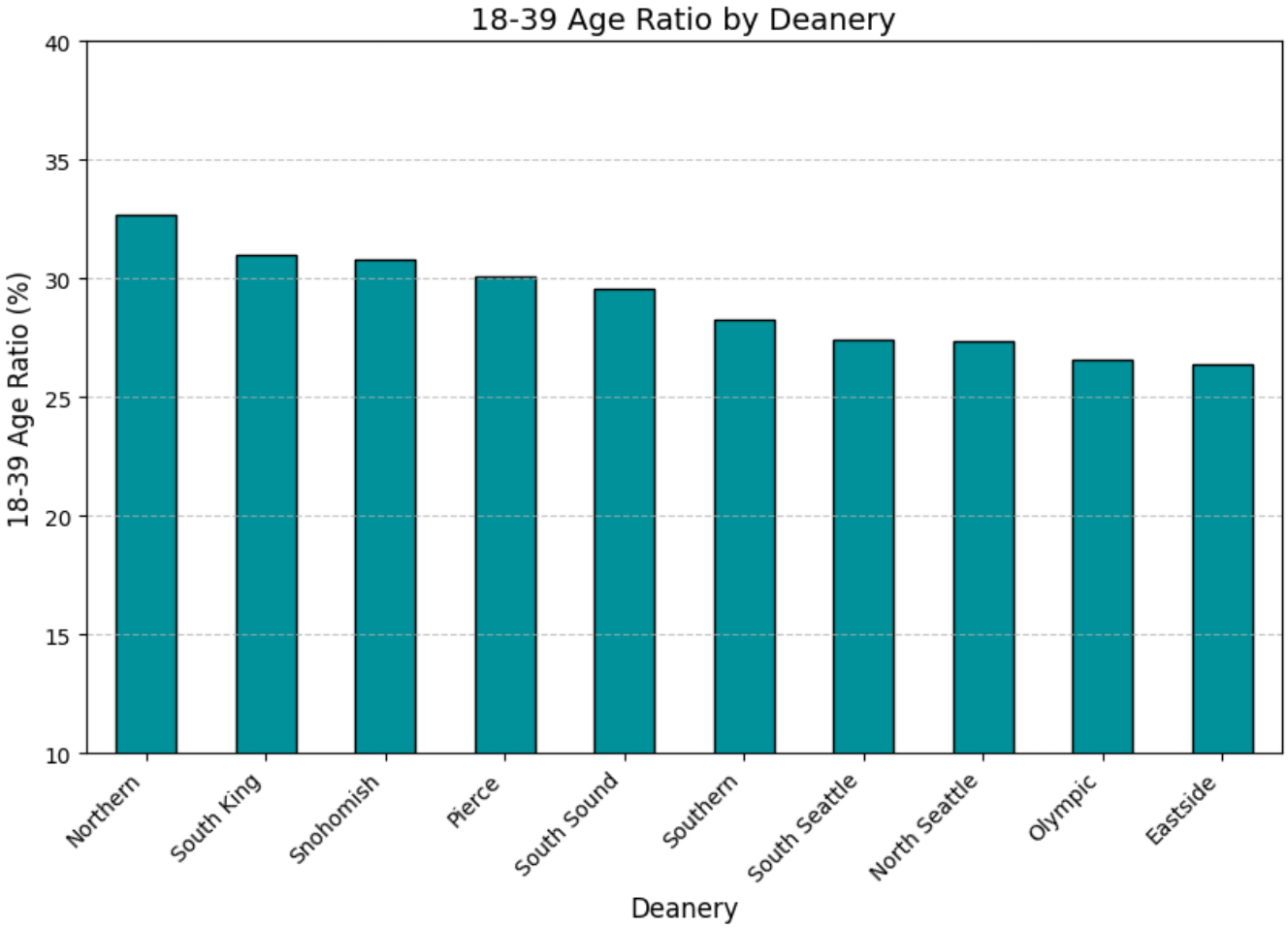


Figure 8. Proportion of Young Adults Across Deaneries

This bar chart compares the proportion of registered young adults across different deaneries. The data reveals noticeable regional differences. The Northern Deanery has the highest percentage of young adults at 33%, while the Eastside Deanery has the lowest at 26%. This suggests that some deaneries may have a stronger young adult presence than others.

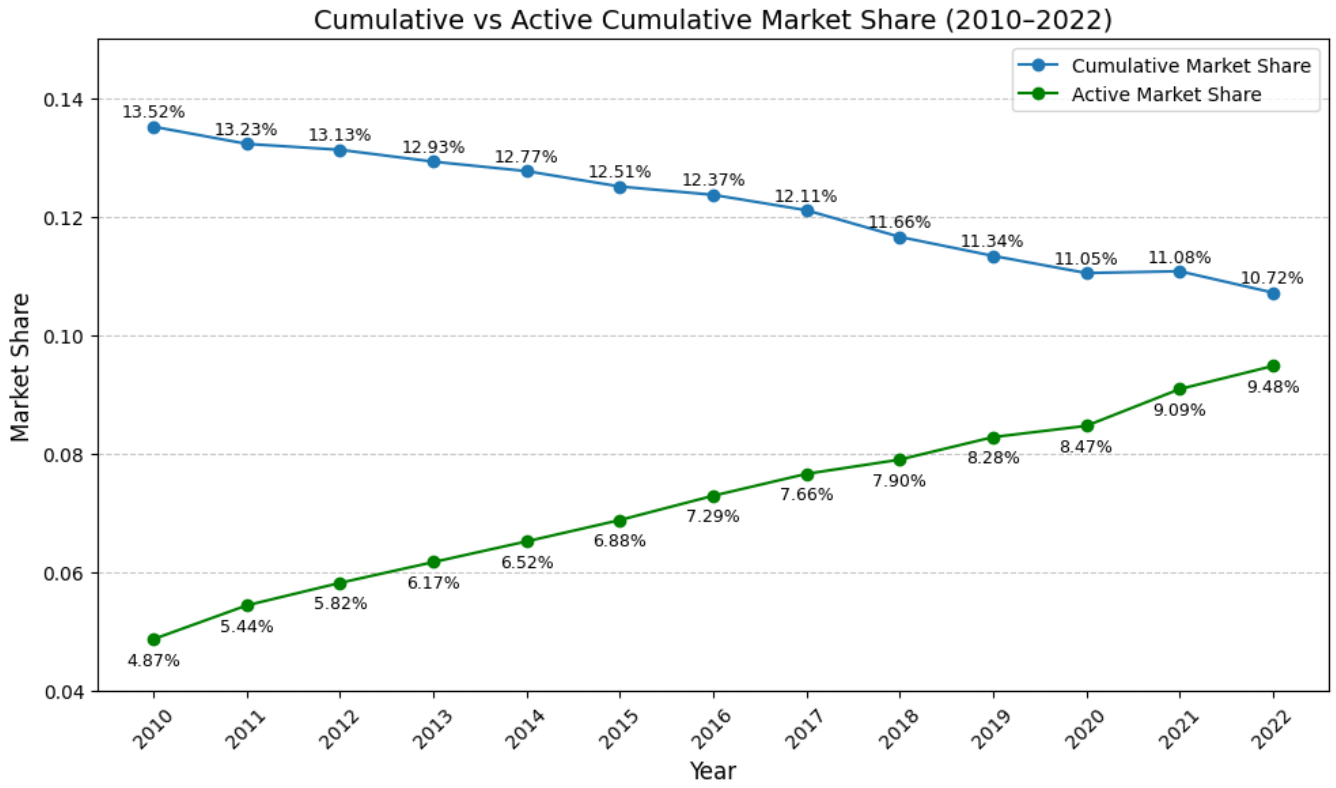
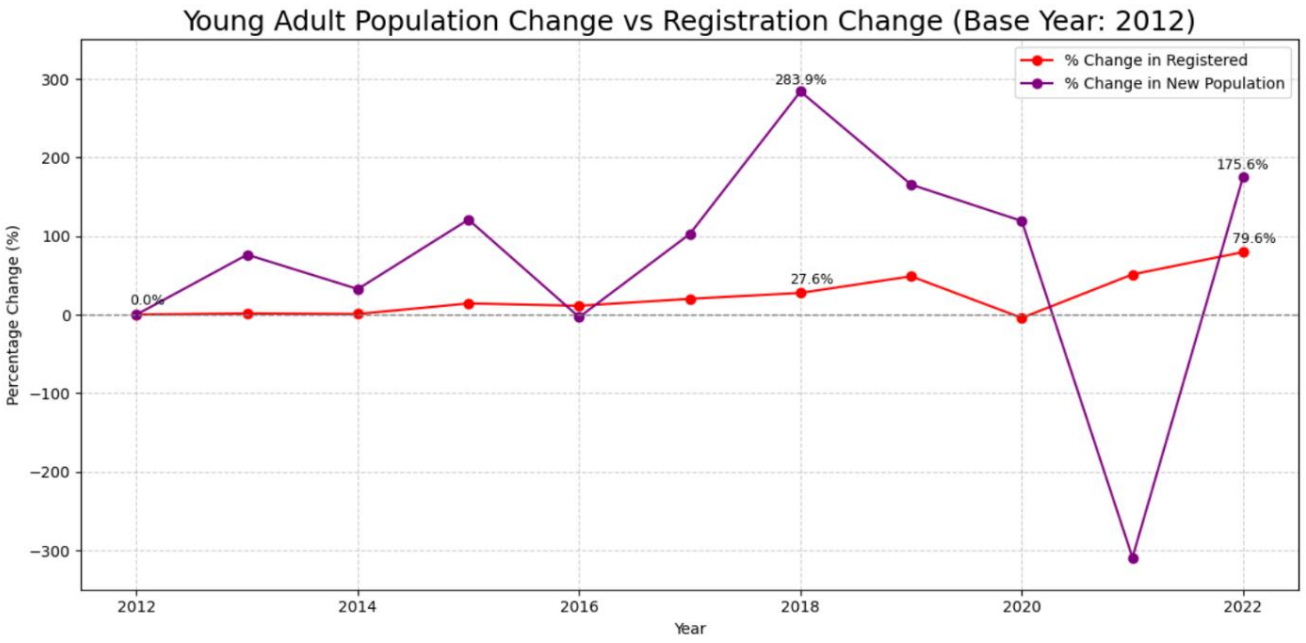


Figure 9. Trends in Young Adult Market Share: Cumulative vs. Active

This chart compares the cumulative and active market share of young adults from 2010 to 2022. Market share is calculated as the number of young adult parishioners divided by the total young adult population in Western Washington.

The blue line represents cumulative market share, including individuals aged 18–39 in a given year who had registered at any point by 2025, regardless of when they joined. The green line shows an active market share, counting only those aged 18–39 who were registered in that specific year. These two perspectives provide a more complete picture by capturing both long-term connection and current participation.



$$\text{Registered Change (\%)} = \frac{\text{Newly Registered in Given Year} - \text{Newly Registered in 2012}}{\text{Newly Registered in 2012}}$$

$$\text{New Population Change (\%)} = \frac{\text{Population Growth in Given Year} - \text{Population Growth in 2012}}{\text{Population Growth in 2012}}$$

Figure 10. Trends in Young Adult Population vs. Registration Growth (2012 Base Year)

This chart compares the percentage change in registered young adults with the change in the young adult population, using 2012 as the baseline year. The red line shows the year-over-year percentage change in the number of newly registered young adults in parishes, while the purple line indicates the percentage change in the new young adult population each year. Significant surges in the young adult population occurred in 2018 (+284%) and 2022 (+176%).

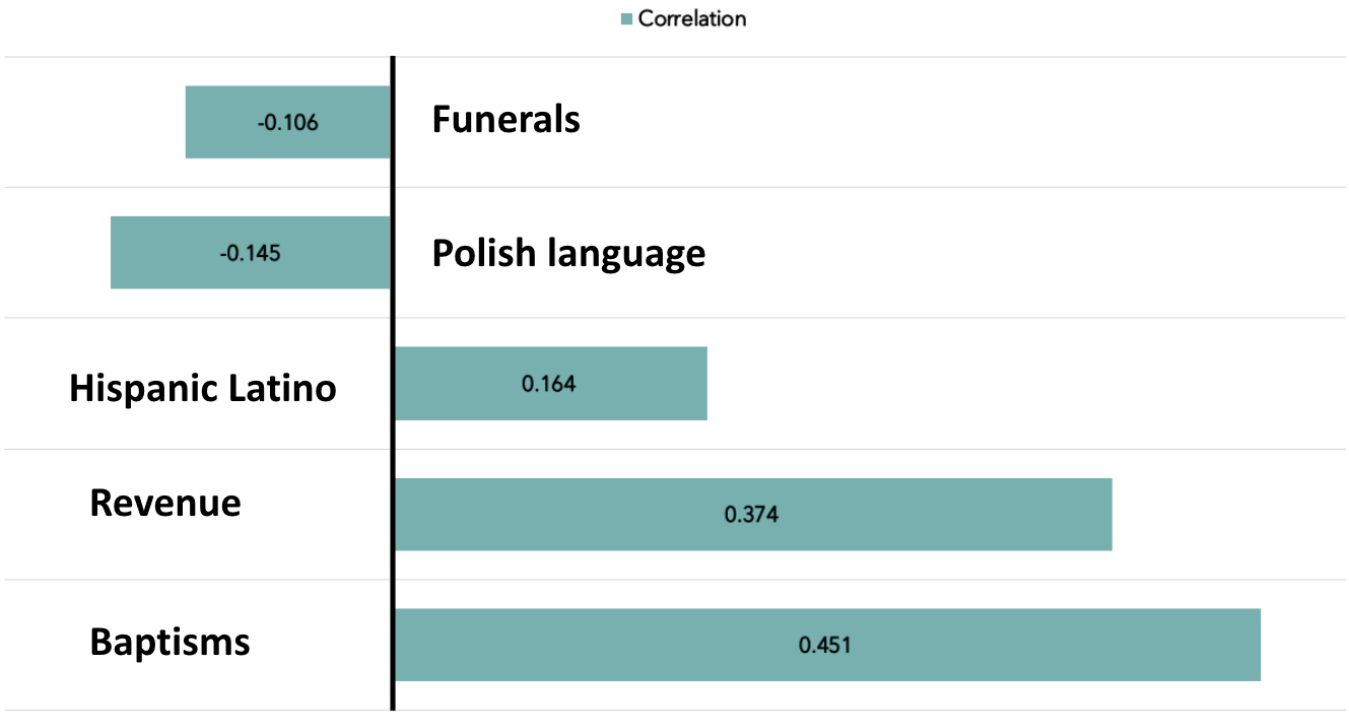


Figure 11. Impact of key features of young adult participation rate

This bar chart displays the correlation between key parish features and the young adult participation rate. Baptisms, parish revenue, and count of Hispanic/Latino parishioners are positively correlated with young adult participation, while Polish language use and funeral counts are negatively correlated.

Variables	Lasso Coefficient	OLS p-value	Correlation
Average number of baptisms	0.033	0.082	0.451
Average parish revenue	0.013	0.345	0.374
Average count of Hispanic/Latino parishioners	0.012	0.073	0.164
Average count of Polish language preference	-0.010	0.089	-0.145
Average number of funerals	-0.011	0.117	-0.106

Table 1. Statistical Summary of Feature Relationships with Young Adult Participation

This table summarizes the strength and direction of association between selected parish-level features and the proportion of young adults among registered parishioners. Results are based on correlation analysis, Lasso regression coefficients, and OLS p-values.

4.4 Parish-Level Impacts of Language and Cultural Services

This bar chart shows the top 5 variables that differ most significantly between "Other Languages Mass offering" parishes and English Mass offering parish, based on T-test results. The y-axis shows -log(p-values), where taller bars mean more significant differences.

Variable	P-Value	Observation
Average confirmations	6.5503e-11	Strong difference
Spanish Language count	7.0114e-11	High Spanish-language presence
Average baptisms	1.0285e-10	Significantly higher or lower
Average communions	2.1741e-10	Notably different
Average mass attendance	1.6001e-08	Less significant but still notable

Table 2. Key Differences Between English-Only and Other Language Parishes

4.5 Impact of Archdiocesan Structural Changes

The analysis focused on two types of structural level actions within the Archdiocese: from Mission to Parish and from Parish to Mission. Each action type revealed different patterns based on key variables identified through T-test results.

Variable	P-Value	Observation
Spanish language count	0.0063	Statistically significant difference
Average funerals	0.0197	Moderately significant difference
Other languages count	0.0204	Moderately significant difference
Single count	0.045050	Moderately significant difference
Avg communions	0.093280	Not statistically significant

Table 3. Key features of conversion of Mission to Parish vs No conversion Parish

This bar chart highlights the top five variables that most significantly distinguish that transition from Mission to Parish compared to those that remained unchanged. Taller bars indicate greater statistical significance based on -log(p-values). Parishes that made this transition tended to offer more Spanish-language Masses, had higher funeral counts, and provided services in other languages. The transition also had more single individuals and a higher number of communions on average.

4.6 Key Variables Influencing Cultural Community Growth

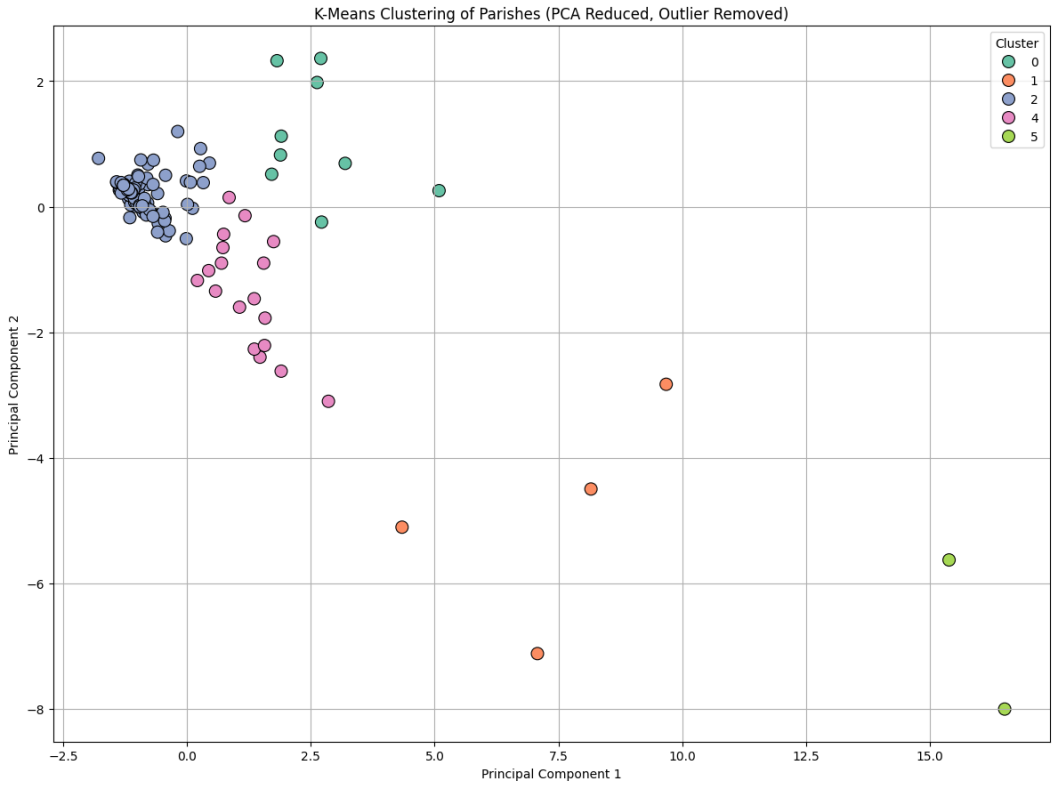


Figure 12. PCA Clusters with K-Means

This scatter plot illustrates the results of K-Means clustering applied to parishes, with their features reduced to two principal components using PCA. Each color on the plot represents a different cluster, labeled from 0 to 5. The axes correspond to Principal Component 1 and Principal Component 2, which together capture the most significant variance in the dataset. Principal Component 1 explains about 23% of the variation and is primarily driven by African American count, White Non-Hispanic count, and populations that are single or divorced, along with the Korean population. This component may reflect demographic diversity and household composition across parishes. Principal Component 2 accounts for another 10% of the variation and is associated with higher counts of Caribbean, Filipino, Japanese, and Southeast Asian populations. It is negatively influenced by the married population, suggesting that parishes with high PC2 values tend to be culturally diverse and have fewer married individuals. These principal components help summarize complex demographic patterns into two dimensions, allowing us to better understand what distinguishes one parish from another. The visualization groups parishes based on shared characteristics, revealing distinct clusters that highlight patterns of similarity in demographics, engagement, or other key attributes. Additionally, a few parishes appear as outliers, indicating unique profiles that differ markedly from the main clusters potentially due to distinctive cultural, geographic, or pastoral dynamics.

5. Discussion

5.1 Cultural Communities within the Archdiocese

This section explores trends, impacts, and implications of cultural communities across parishes.

5.1.1 Future Trend Prediction

An Arima Time Series Model was implemented to estimate the future parish population, which showed that there would be a deciding trend in the parish population in the future. To evaluate the model performance data has been adjusted to let the model estimate the actual last two data points i.e. (parish population of 2023 and 2024). The model has provided reliable estimates with very an error percentage of 1% suggesting that the future parish population would follow a similar declining trend. It would be presumptuous to assume that the future population trend will also follow or mirror the model's prediction as certain factors like faith in religion or potential demography change are not accounted for by the model.

5.1.2 Key Variables Influencing Cultural Community Growth

The results of K-Means clustering, informed by principal component analysis (PCA), provide significant insights into the demographic and cultural composition of parishes across the Archdiocese of Seattle. The majority of parishes group into three dominant clusters, suggesting shared characteristics particularly in terms of ethnic background and marital demographics. This structural consistency points to the viability of unified pastoral strategies and resource-sharing initiatives, which can strengthen participation and operational efficiency within these broadly similar communities.

In contrast, a smaller number of parishes fall into more isolated clusters, visually and analytically distinct in the PCA space. These parishes are characterized by higher proportions of specific cultural groups such as Caribbean, Middle Eastern, Filipino, and Southeast Asian communities. Their demographic uniqueness suggests that one-size-fits-all strategies may be ineffective. Instead, these communities would benefit from more customized approaches, including tailored language Masses, targeted outreach efforts, and ministries that are culturally responsive and community specific.

Metric	Value
African American/Black Count per Mass Attendance	0.283150
White Non-Hispanic Count per Mass Attendance	0.270955
Single Count per Mass Attendance	0.270005
Korean Count per Mass Attendance	0.262753
Divorced Count per Mass Attendance	0.241718
Married Count per Mass Attendance	0.237201

Table 4. PC1 components

The underlying principal components further support these findings. PC1 captures a wide range of racial and marital diversity, with strong loadings on variables such as African American, White Non-Hispanic, and marital status per mass attendee. PC2, meanwhile, distinguishes parishes with higher proportions of Caribbean, Southeast Asian, and Middle Eastern populations, indicating a different axis of cultural segmentation. Together, these components highlight the multidimensional nature of parish identity, encompassing both cultural heritage and life-stage markers.

Metric	Value
Caribbean Count per Mass Attendance	0.380721
Southeast Asian (Other) Count per Mass Attendance	0.364348
Filipino Count per Mass Attendance	0.340950
Middle Eastern Count per Mass Attendance	0.334427
Japanese Count per Mass Attendance	0.315247
Married Count per Mass Attendance	0.255660

Table 5. PC2 components

From a strategic planning perspective, these insights serve as a roadmap for more data-driven pastoral decision-making. The concentration of parish within a few major clusters indicates where integrated and standardized initiatives can drive participation and cohesion. Meanwhile, the distinct profiles of smaller clusters call for flexible planning and tailored programming to ensure inclusivity and equity. By responding to these nuanced patterns with targeted interventions, the archdiocese can better serve its diverse Catholic population and more effectively support the goals of the archdiocese.

5.1.3 Parish-Level Impacts of Language and Cultural Services

To examine the influence of language-specific Masses on parish-level outcomes, parishes were divided into two groups: those offering Mass in a language other than English (Group A) and those that did not (Group B). A total of 52 parishes were identified as offering non-English Masses, based on available records. To assess the differences between these two groups, two independent t-tests were conducted across various parish-level numerical indicators, including sacramental activity, attendance, demographic representation, and financial metrics.

The results revealed statistically significant differences across several key variables. Parishes offering Mass in another language had notably higher levels of sacramental participation. Specifically, the average number of confirmations, baptisms, and communions were significantly higher in these parishes, with p-values less than 0.001. For instance, the average confirmations variable yielded a t-statistic of 6.85 with a p-value of approximately 1.3e-10, indicating a strong association between multilingual services and higher participation in this sacrament. Similarly, mass attendance was also significantly higher in these parishes, further reinforcing the idea that language-specific services enhance engagement.

Interestingly, while sacramental participation and attendance showed strong associations, financial indicators such as average income, revenue, and expenses did not differ significantly between the two groups. This suggests that offering Mass in other languages may not directly lead to increased

financial contributions, but it does correlate with greater community involvement in spiritual and sacramental life. Additionally, the average number of households was significantly higher in parishes with non-English Masses, indicating that such parishes may also be serving larger or more active communities.

To validate whether this effect was not just different, but positively higher for Group A, a one-tailed t-test was conducted to specifically test the hypothesis that Group A performs better than Group B. The results supported this hypothesis that all the sacramental metrics, including baptisms, communions, confirmations, marriages, funerals, and average household size, were significantly higher in Group A. Spanish-speaking populations were significantly more prominent in parishes that offered non-English Masses, as indicated by the high t-statistic and low p-value for the Spanish language count. This trend is likely driven by the fact that the Archdiocese of Seattle offers a greater number of Masses in Spanish compared to other languages, reflecting both the size of the Spanish-speaking Catholic population and institutional prioritization. In contrast, Vietnamese and Korean community representation did not show statistically significant differences between the two groups. This suggests that their participation may be shaped more by geographic distribution or community-specific factors than by the presence of language-specific Masses alone.

The presence of non-English Masses appears to be a strong driver of sacramental participation and community engagement. While not directly tied to financial outcomes, these services play a critical role in fostering inclusivity and ensuring active participation among culturally diverse parishioners.

5.1.4 Impact of Archdiocesan Structural Changes

To evaluate the effects of structural changes within the archdiocese, statistical tests were conducted for two groups of parishes. The first group included parishes that transitioned from missions to fully established parishes, while the second group included parishes that were formerly parishes but later became missions. Both groups were compared against parishes that had not undergone any change in their structural status.

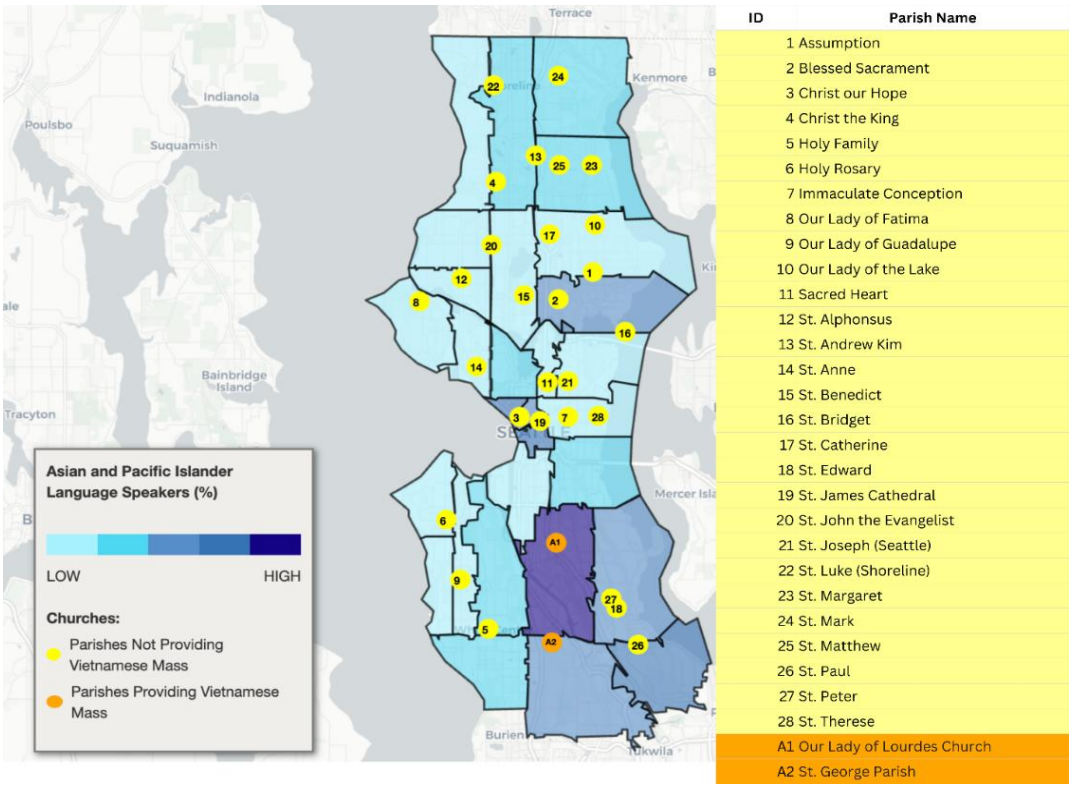
In both transition cases, the results showed a consistent trend. Parishes that experienced structural change had significantly higher levels of sacramental activity. This included higher averages for baptisms, communions, confirmations, funerals, marriages, and Mass attendance. Notably, this pattern held true regardless of the direction of the transition, whether from mission to parish or from parish to mission. These findings suggest that parishes involved in transitions tend to serve more actively engaged communities. Demographic indicators such as counts of male and female parishioners, children, young adults, adults, and married individuals were also higher. In addition, these parishes showed greater language diversity, especially with respect to Spanish-speaking and other non-English groups.

In contrast, financial indicators including average income, revenue, and expenses did not show significant differences between transitioned and non-transitioned parishes. Similarly, ethnic group representation across communities such as Vietnamese, Filipino, African American, and others

remained largely unchanged. These results suggest that while structural changes are linked to stronger community participation and broader inclusion in liturgical life, they are not necessarily associated with financial performance or major demographic shifts. The patterns observed may reflect existing community strengths or decisions based on pastoral planning rather than participation alone.

5.1.5 Spotlight on Key Cultural Communities

The Spanish and Vietnamese communities are experiencing steady population growth. Parishes that offer masses in this language have also seen growth in registrations. The two maps below illustrate the density of Spanish and Vietnamese populations in Seattle, overlaid on them are the local churches. The churches that offer these language-specific masses are categorized differently from the remaining churches (shown in a different color). The reason for creating this map is to identify areas with a high density of specific populations and to check whether certain churches are not offering language-specific masses despite the demand, so this can be brought to the attention of the organization.

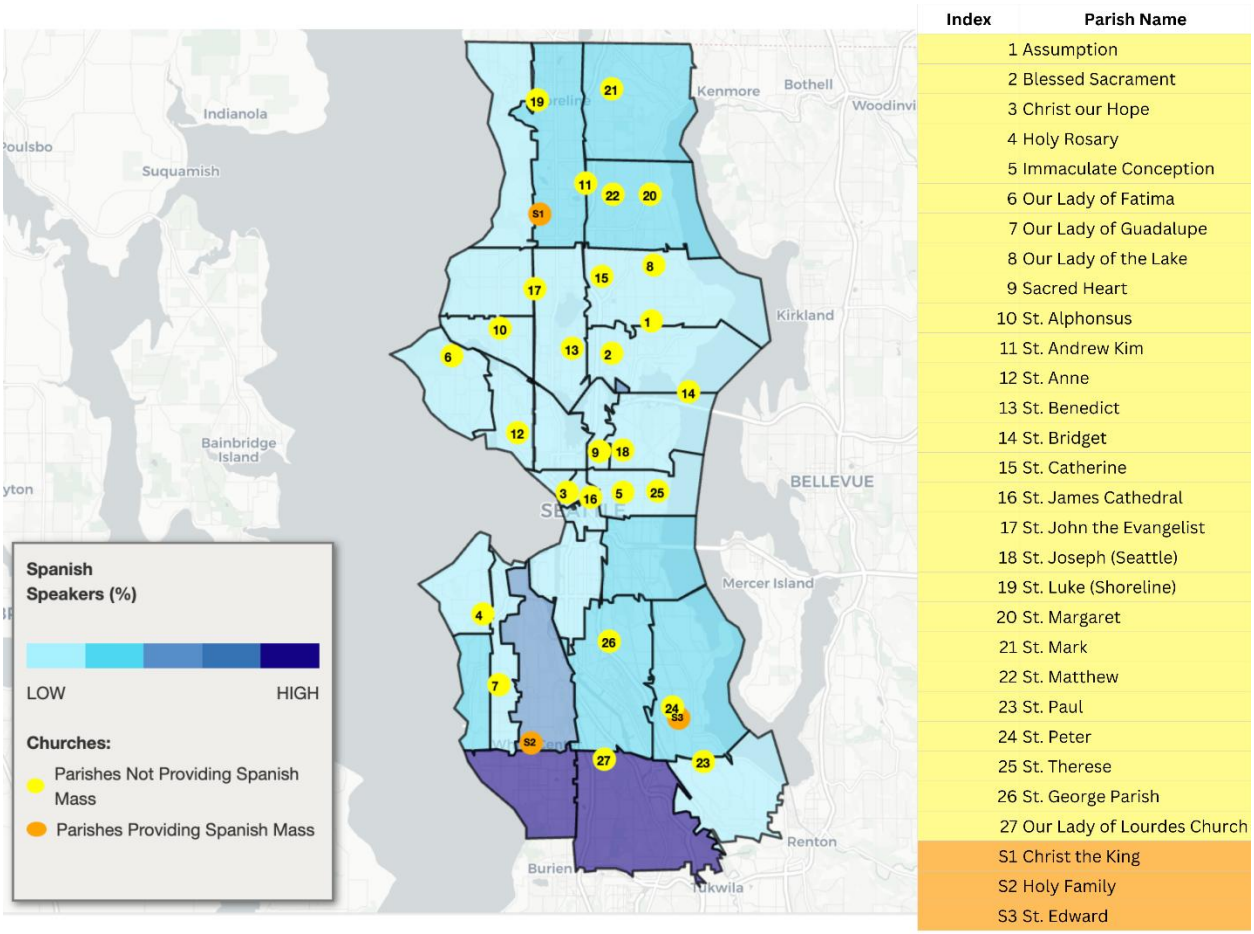


ID	Parish Name
1	Assumption
2	Blessed Sacrament
3	Christ our Hope
4	Christ the King
5	Holy Family
6	Holy Rosary
7	Immaculate Conception
8	Our Lady of Fatima
9	Our Lady of Guadalupe
10	Our Lady of the Lake
11	Sacred Heart
12	St. Alphonsus
13	St. Andrew Kim
14	St. Anne
15	St. Benedict
16	St. Bridget
17	St. Catherine
18	St. Edward
19	St. James Cathedral
20	St. John the Evangelist
21	St. Joseph (Seattle)
22	St. Luke (Shoreline)
23	St. Margaret
24	St. Mark
25	St. Matthew
26	St. Paul
27	St. Peter
28	St. Therese
A1	Our Lady of Lourdes Church
A2	St. George Parish

Figure 13. Availability of Vietnamese Masses Relative to Language Speakers in Parishes

The accurate Vietnamese population data in Seattle is not available, So Asian and Pacific Islander Language speakers' data has been acquired as a proxy. Since Vietnamese is one of the dominant Asian languages in Seattle, using it as a proxy would not distort the result that are being acquired. Both A1 and A2 (Vietnamese-Mass offering parishes) are in locations where there is a high Vietnamese Density. This indicates that the organization is being responsive to cultural changes.

However, certain areas like the U-District have a higher population density and no churches are offering masses in Vietnamese, including masses in north Seattle will be beneficial for the organization to better serve these communities and align more closely with local demographic needs.



Index	Parish Name
1	Assumption
2	Blessed Sacrament
3	Christ our Hope
4	Holy Rosary
5	Immaculate Conception
6	Our Lady of Fatima
7	Our Lady of Guadalupe
8	Our Lady of the Lake
9	Sacred Heart
10	St. Alphonsus
11	St. Andrew Kim
12	St. Anne
13	St. Benedict
14	St. Bridget
15	St. Catherine
16	St. James Cathedral
17	St. John the Evangelist
18	St. Joseph (Seattle)
19	St. Luke (Shoreline)
20	St. Margaret
21	St. Mark
22	St. Matthew
23	St. Paul
24	St. Peter
25	St. Therese
26	St. George Parish
27	Our Lady of Lourdes Church
S1	Christ the King
S2	Holy Family
S3	St. Edward

Figure 14. Availability of Spanish Masses Relative to Language Speakers in Parishes

The organization has been responsive in including masses in high Spanish-density areas. This shows again that the organization is responsive to cultural changes. The masses are offered in churches where there is a high density of Spanish-speaking population especially in south Seattle. To balance the masses offered across Seattle one church in the north is also offering Spanish mass as shown in the map above.

5.2 Young Adult Participation Trends

5.2.1 Current Market Share and Active Market Share

To analyze how young adults have connected with parish communities over time, this analysis focuses on two kinds of participation. The first is active market share, which tells us how many young adults were already registered in a parish during that same year. The second measure is called cumulative market share. This tells us how many young adults from each year eventually joined the parish, even if they didn't join right away.

In recent years, active market share has been going up. This means that a higher percentage of young adults were already registered with a parish during the same year they were counted. In other words, it shows how many young people the parish was actively connected to in that year. As a result, recent outreach efforts have been successful in attracting more young people and encouraging them to participate earlier.

At the same time, cumulative market share helps us understand long-term engagement. Some people may not register at first, but they still join later on. By examining the complete data through 2025, it becomes clear that many young adults do become part of the parish community over time, even if they did not join immediately. A good example is the group from 2013. Only a small number of young adults from that year were already registered, but by 2025, a much larger portion had joined. This shows that long-term connection is possible, and that some people take more time to become involved.

By looking at both active and cumulative market share together, we get a fuller picture. We can see who is participating now, and we can also understand how young adults stay connected over time. These insights can help the parish create better outreach plans, by supporting both short-term involvement and long-term relationships.

5.2.2 Trends in Young Adult Population vs. Registration Growth

The number of young adults registering has been steadily going up since 2012, which shows that current outreach efforts are working well overall. However, in years with sharp increases in the young adult population, such as 2018 and 2022, registration growth did not keep pace. This indicates missed opportunities for engagement during periods of population expansion.
To address this gap, parishes should consider developing proactive outreach strategies in years when demographic data predict a rise in the young adult population. Acting early can help make sure that more people register when the population increases.

A notable shift in the trend occurred in 2021. That year, the young adult population dropped sharply, probably because of the COVID-19 pandemic. Many young adults may have moved away for remote work or returned home to live with their families. Even though the population dropped, the number of newly registered young adults still went up. This divergence suggests that parish outreach efforts remained effective even under challenging circumstances. It may also indicate that

those who remained in the region were more likely to seek connection and stability through the church during a time of uncertainty.

5.2.3 Variables Influencing Young Adult Engagement

There are clear differences in young adult participation across deaneries. Northern and South King show the highest ratios, while Eastside and Olympic have the lowest. These differences may reflect variations in parish activities, community programs, demographic composition, and available resources, or they may simply be influenced by the fact that urban areas tend to have a higher concentration of young adults in the general population.

However, our analysis focuses on the proportion of young adults within each parish population, rather than the overall population in each geographic area. While this limits broader demographic interpretation, it still provides meaningful insights into the internal age structure of parishes and helps identify parish-level characteristics that may support or hinder young adult engagement.

To better understand these patterns, we examined the relationship between several parish-level features and the ratio of young adults to the total parish population using correlation analysis, Lasso regression, and OLS regression. The features with the strongest positive and negative associations include baptisms, parish revenue, Hispanic/Latino parishioners, Polish language use, and funeral counts.

Among these, baptisms showed the strongest positive correlation with young adult participation (correlation = 0.451), and its Lasso coefficient was the highest (0.033). This suggests that parishes with more baptisms tend to have greater young adult engagement, likely because baptism often represents a key entry point for young individuals or families returning to the church. Parish revenue (correlation = 0.374) and Hispanic/Latino presence (correlation = 0.164) also showed positive associations. The connection between revenue and participation may reflect broader parish capacity, while the association with Hispanic/Latino parishioners may be partly explained by the younger age profile of this demographic group, as well as strong family and religious traditions that encourage church involvement.

In contrast, Polish language use and funeral counts both had negative correlations with young adult participation and were selected by the Lasso model. These features are commonly linked to older or more traditional parish populations, which may help explain lower engagement among younger members.

Notably, while the OLS p-values for some features such as parish revenue and funeral counts were not statistically significant, their inclusion in the Lasso model suggests they still provide explanatory value in a regularized regression context. This shows the usefulness of Lasso in identifying variables that may not show strong individual associations but still contribute meaningfully when considered together with other factors.

These findings provide a foundation for designing outreach strategies that are tailored to the demographic, cultural, and structural characteristics of each region.

5.3. Actionable Insights

Based on the data-driven insights uncovered in this project, several actionable strategies emerge that can immediately inform pastoral planning and community outreach across the Archdiocese of Seattle.

First, there is a clear need to expand Spanish-language Masses in high-density areas where current offerings are insufficient. GIS mapping revealed that while the Archdiocese has responded well in some regions, such as South Seattle, there are still notable gaps, particularly in North Seattle and the University District. Addressing this mismatch between community demographics and available services will significantly enhance engagement among Hispanic populations.

Second, the data shows that young adult participation is increasing, especially in deaneries like Northern and South King. However, registration spikes often lag with population surges. To close this gap, the Archdiocese should implement data-informed outreach plans that align with forecasted population growth. For example, leveraging ARIMA projections to anticipate demographic shifts could help target young adults during high-growth years like those seen in 2018 and 2022.

Third, the analysis revealed that structural changes at the parish level, particularly transitions between missions and parishes, correlate with higher sacramental engagement. These findings suggest that parish conversions can serve as a powerful tool to revitalize stagnant communities. Strategic use of this mechanism could increase attendance and participation in sacraments like Baptism and Communion, especially in areas currently experiencing a decline.

Fourth, clustering analysis (PCA + K-Means) identified several parishes with unique cultural compositions, such as Vietnamese, Filipino, and Caribbean communities. These parishes should be cultivated as cultural epicenters by tailoring pastoral strategies, offering culturally specific Masses, and investing in community-based leadership and events. Doing so would reinforce their identity and deepen their role within the larger faith network.

Fifth, to manage and scale these efforts effectively, the Archdiocese should implement a centralized ministry dashboard, using platforms like PowerBI. This system would consolidate key parish-level metrics, mass attendance, sacraments, demographic shifts into a real-time, visual decision-making tool that enables parish leaders and planners to act swiftly and with precision.

Lastly, this project's forecasting tools and statistical methods, such as the ARIMA model and T-tests, should be formalized into a predictive planning framework. By embedding these tools into the Archdiocese's annual planning cycle, leadership can proactively identify areas at risk for decline and uncover new opportunities for outreach and growth.

6. Conclusion

This project provides a data-driven framework to analyze the demographic and engagement trends within the Archdiocese of Seattle, with a particular focus on cultural communities and young adults. By leveraging statistical modeling, machine learning techniques, and data visualization, we aim to generate actionable insights that will support strategic pastoral planning under the *Partners in the Gospel* initiative. Our approach integrates multiple analytical methods: ARIMA for forecasting parish participation trends, Principal Component Analysis (PCA) for identifying key demographic drivers, and K-means clustering to segment parishes based on engagement patterns. PCA revealed two primary components: PC1, which captures racial and marital diversity with strong loadings on variables such as African American, White Non-Hispanic populations, and marital status per Mass attendee; and PC2, which distinguishes parishes with high proportions of Caribbean, Southeast Asian, and Middle Eastern populations highlighting cultural segmentation. These models, supported by SQL databases, provide a scalable and interpretable solution to assess community growth and participation.

The outcomes of this study will help the archdiocese allocate resources more effectively, tailor outreach efforts, and enhance engagement strategies for diverse cultural communities and young adults. Additionally, these insights will inform policy decisions related to parish mergers, new language ministries, and strategic consolidations. The long-term impact of this work lies in its ability to offer an adaptable, data-driven planning tool that can continuously support the Archdiocese's mission of fostering a vibrant and inclusive Catholic community. By combining historical data analysis with predictive modeling, this project ensures that future pastoral strategies are grounded in real-world demographic shifts. Ultimately, our findings will empower Church leadership with the knowledge necessary to create sustainable, thriving parish communities across the Archdiocese of Seattle.

References

1. Nortey, Justin, et al. “10 Facts about U.S. Catholics.” Pew Research Center, 4 Mar. 2025. < https://www.pewresearch.org/short-reads/2025/03/04/10-facts-about-us-catholics/>
2. “People in Washington: Religious Landscape Study.” *Pew Research Center*, Accessed 5 Mar. 2025. <www.pewresearch.org/religious-landscape-study/state/washington/>
3. “Archdiocese of Seattle” Pastoral Plan - Archdiocese of Seattle. (Vision Section), Accessed 24 Jan. 2025. < https://archseattle.org/about-the-archdiocese-of-seattle/archbishop-etienne/pastoralplanning/pastoral-plan/>
4. censusdis. *censusdis-tutorial-2024*. GitHub, 31 Oct. 2024, https://github.com/censusdis/censusdis-tutorial-2024.
5. Nadeem, Reem. “Modeling the Future of Religion in America.” *Pew Research Center*, Pew Research Center, 13 Sept. 2022, www.pewresearch.org/religion/2022/09/13/modeling-the-future-of-religion-in-america/
6. “SQLite.” *Wikipedia*, Accessed 20 Feb. 2025, <https://en.wikipedia.org/wiki/SQLite.>
7. “IPUMS.” *Ipums.org*, 2019, <www.ipums.org/>.
8. “Autoregressive Integrated Moving Average (ARIMA).” Investopedia, 31 July 2024, www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp Ravelo,
9. Ciera. “The Stationary Data Assumption in Time Series Analysis.” Statistics Solutions, 31 Jan. 2023, www.statisticssolutions.com/stationary-data-assumption-in-time-series-analysis/
10. Works CitedCFI Team. “Autocorrelation.” Corporate Finance Institute, 3 Nov. 2022, corporatefinanceinstitute.com/resources/data-science/autocorrelation/Hayes, Adam.
11. Nau, Robert. “Introduction to ARIMA Models.” *Duke.edu*, 2019, people.duke.edu/~rnau/411arim.html.
12. Keboola. “PCA in Machine Learning: A Comprehensive Guide.” Accessed 22 Feb. 2025, < https://www.keboola.com/blog/pca-machine-learning. >
13. Analytics Vidhya. “Comprehensive Guide to K-Means Clustering.” *Analytics Vidhya*, Accessed 22 Feb. 2025, <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/. >
14. IBM “What Is Lasso Regression?” *IBM*, 16 Apr. 2025, <www.ibm.com/think/topics/lasso-regression#:~:text=Lasso%20regression%20is%20a%20regularization,regularization%20for%20linear%20regression%20models>
15. “Understanding Ordinary Least Squares (OLS) Regression.” *Built In*, <builtin.com/data-science/ols-regression>. Accessed 20 Apr. 2025.
16. Datagy.io. "How to Perform T-Tests in Python (One- and Two-Sample)." Datagy, 2023, https://datagy.io/t-test-python/.