Introduction

According to 2021 Census data, 23 different languages are commonly spoken in San José and 60% of San José residents speak a language other than English at home. Many individuals living in the City whose primary language is not English may have limited English proficiency (LEP). Based on current estimates, 23% of San José residents speak English "less than very well." The two most popular non-English languages spoken in San José are Spanish and Vietnamese. Spanish-speaking residents make up 10.3% of San José's total population and Vietnamese speaking residents make up 6.5% of San José's total population[1]. Based on this, the City of San José decided to increase access through these two languages on all its platforms and mode of interaction with the residents.

Increasing language access involves both having high-quality resources available in different languages and ensuring that populations that need them are made aware of these resources. This showed to be the case with the San José 311 (SJ311) platform. In 2020, the City integrated Spanish and Vietnamese translation into the SJ311 platform to allow residents to create service requests in those two languages. This was a multiple-step process that required extensive data-driven methods and decision making. This data story outlines the work of SJ311 in creating a Spanish and Vietnamese translation tool and strategically outreaching to target populations.

Addition of Google Language Translation Services

Since its launch in 2017, SJ311 services have only been available in English. To expand access, SJ311 expanded its language access to Spanish and Vietnamese residents.

An initial solution that was evaluated was an out of the box translation service on the platform. Two out of three Spanish speakers and five out of seven Vietnamese speakers preferred the simplified English over the translation. In addition, the out of the box translation was found to be offensive in some cases. As a result, the City started to explore additional translation solutions.

The challenge was how to provide the best translation services for users. The choice was between human and machine learning (ML) translation. Human translation is slow and costly but high quality. Machine translation is fast and free but low quality. As a middle ground, City conducted a pilot with Google's Auto ML machine learning translation services to provide a customized translation solution built on top of the Google Free Translation. The Auto ML model was trained on words and phrases (training pairs) specific to the City and 311 context that reflected the regional dialects of San José's Spanish and Vietnamese speaking community. Spanish speaking residents preferred a Mexican dialect and Vietnamese speaking residents preferred pre-1975 Southern dialect.

The introduction of Auto ML brought about various assumptions regarding its process. Initially, it was believed that translation functioned uniformly for all types of text and that a minimum of 10,000 phrase pairs were necessary to achieve top-notch translation quality. Nevertheless, subsequent findings demonstrated that these assumptions were unfounded.

---

[1] US Census Bureau, 2010-2014 American Community Survey

The City-trained model revealed spelling errors and Vietnamese was harder to translate. ML models were not able to translate misspelled words, making it challenging to interpret what users were inputting. In addition, cultural contexts and dialect made it challenging to translate certain words such as fire hydrant and graffiti.

Significantly less than 10,000 phrase pairs were used in both language models. 1,178 phrase pairs were used in the Spanish model and 2,049 pairs were used in the Vietnamese model. Though these limited phrase pairs were sufficient to train the model, Google's base model learns from Auto ML every 18 months. This means the City needs to add additional training pairs to get the most out of the machine learning model.

To assess the quality of the City-trained ML translation tool, human evaluators were used. Twenty human evaluators rated 100 sentences on a three-point scale (Good, Acceptable, Needs Work). A final rating agreement from three out of four evaluators were needed to pass. If a rating did not pass, a human translator was used to retrain the model. Humans rated the City tool better than Google's free translation by 22-51%.

Translation direction	Free translation tool (Percentage rated by humans as "Acceptable" or better)	City-trained translation (Percentage rated by humans as "Acceptable" or better)	Improvement
English to Spanish	66%	98%	+ 32%
Spanish to English	62%	88%	+ 26%
English to Vietnamese	37%	88%	+ 51%
Vietnamese to English	44%	66%	+ 22%

Current SJ311 Platform

When a user creates a service request (SR) on the 311 platform, the SR content is translated into the user's preferred language. The platform uses a combination of a predefined glossary, Google's free translation, and Google ML model. Content is translated through a three-step process. All static content, such as the SR names and other descriptive text, is first translated using the predefined glossary. Any content not in glossary is translated through Google's free translation. Dynamic content, such as user-provided content in the open text fields, are translated using Google Auto ML model. Requests are then inputted in the user's desired language and translated to English so call agents may work on them. When City staff responds to the request all response content is dynamically translated from English using Auto ML to the user's preferred language and displayed to the user in that language.

Increasing Engagement to Platform

After Spanish and Vietnamese translations were added onto the SJ311 platform, the use of the languages remained low.

The solution was to target outreach to areas with high percentages of Spanish and Vietnamese speaking residents. SJ311 and City Manager's Office of Communications worked on targeted community outreach to promote awareness of SJ311 services.

SJ311 worked with the City's Geographic Information System (GIS) team to create the SJ311 Equity dashboard. The Equity dashboard utilized City Equity Atlas[2] that uses Equity score as a combined score of race and income for identifying geographical areas of high underserved populations in the City. It also utilized the data language score for identifying the high Spanish and Vietnamese speaking areas in the City. The Equity dashboard brought together data from SJ311 usage, equity score, and language score. This helped identify zip codes with larger underserved populations and people with LEP. Areas with equity scores above eight and Language scores four and five were prioritized for outreach.
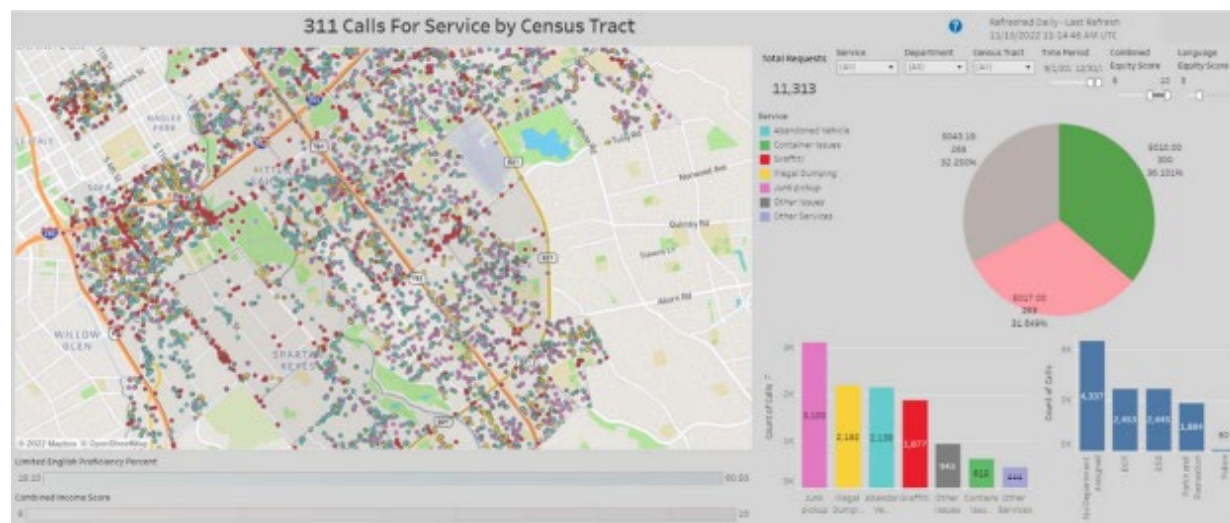

Figure 1 – SJ311 Community Outreach Equity Dashboard

Outreach was approached in two stages. Phase one spanned from April 2022 to September 2022. Phase two spanned from November 2022 to March 2023. Various outreach strategies were used. Low-cost tactics included posters and handouts at public spaces such as libraries and community centers. Other tactics were more costly such as targeted digital ads on Facebook and Google, and direct mail postcards to target locations.

---

[2] Equity Atlas - https://gis.sanjoseca.gov/maps/equityatlas/

	2022									2023				
	Phase 1									Phase 2				
	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Organic social media post	4/14							11/22			2/17			
Nextdoor	4/24		6/2					11/22			2/17			
Poster distrbution to community centers	4/25							11/29						
Google digital ads (trilingual)	4/25-5/16										2/21-3/21			
Poster distrbution to libraries		5/4						11/23						
Direct mail to 95116		5/7												
Email to neighborhood associations		5/17												
Targeted FB ads		5/26-6/10												
Email - Toolkit to SJ311 departments			6/1											
Email to garbage and recycling companies			6/3											
Cardstock handout distrbution to depts and partners			6/8											
Cardstock handout distrbution to community centers			6/22											
Direct mail to 95112										1/3				
Grand Century mall ads (Vietnamese)											2/20 - - - - - - - - - - - - - - 5/24			
Univision radio ads (Spanish)												3/6-3/31		
Billboard "poster" in 95116 - Alum Rock (Spanish)										1/20 - - - - - - - - 3/17				
Billboard "poster" in 95112 - Bayshore (Spanish)											2/20 - - - - - - - - - 4/17			
PRNS summer activity guide														
Neighborhood association presentations														

Figure 2 – Timeline of SJ311 Outreach Tactics

An overall increase in engagement occurred with SJ311 after Phase one of outreach. Between the end of 2021 to the middle of 2022, nearly 60% increase in SRs occurred.
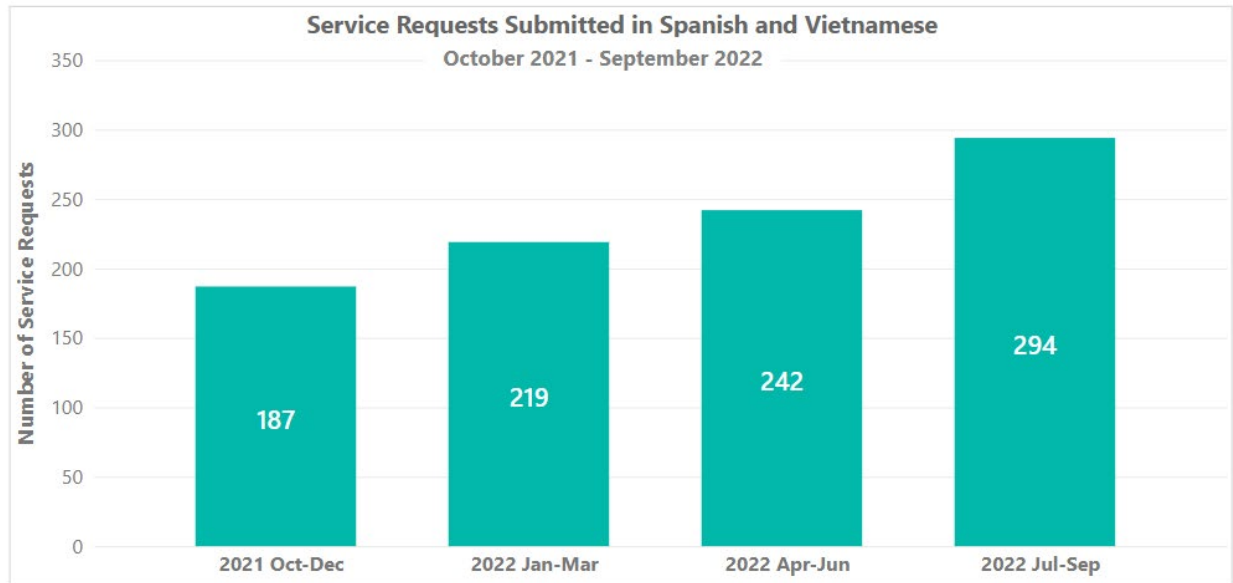

Figure 3 – 311 requests in Spanish and Vietnamese from 2021 to 2022

An increase in calls to the call center from both Spanish and Vietnamese callers from 2021 to 2022 occurred. Due to the high volume of Spanish callers, a virtual agent was added to the call center. Virtual agents were able to handle simpler requests which allowed live agents to handle more complicated calls. From 2021 to 2022, a 132% increase of Spanish calls (live and virtual agents combined) and 21.9% increase in Vietnamese calls occurred.

	CY 2021	CY 2022
Spanish Calls	6,322	5,065
Spanish Virtual Agent		9,859
Vietnamese Calls	1,478	1,803

Figure 4 – Number of Calls in Spanish and Vietnamese for 2021 and 2022

One of the largest outreach initiatives was Google Ads. Google Ads allowed the SJ311 team to monitor website traffic directly through the ads. Google Ads resulted in 1.3+ million impressions and 16,890 clicks to the SJ311 website. Cost per click of the campaign was $0.21 which is less than the government average at $0.81.

Following the Google and Facebook Ad campaigns and the direct mail postcards, an increase in SRs occurred from zip codes with a high percentage of Spanish and Vietnamese speakers. Spanish targeted zip codes including 95116, 95112, and 95127 showed a substantial increase in Spanish SRs; Vietnamese targeted zip codes including 95111, 95112, 95148, and 95136 showed a substantial increase in Vietnamese SRs.
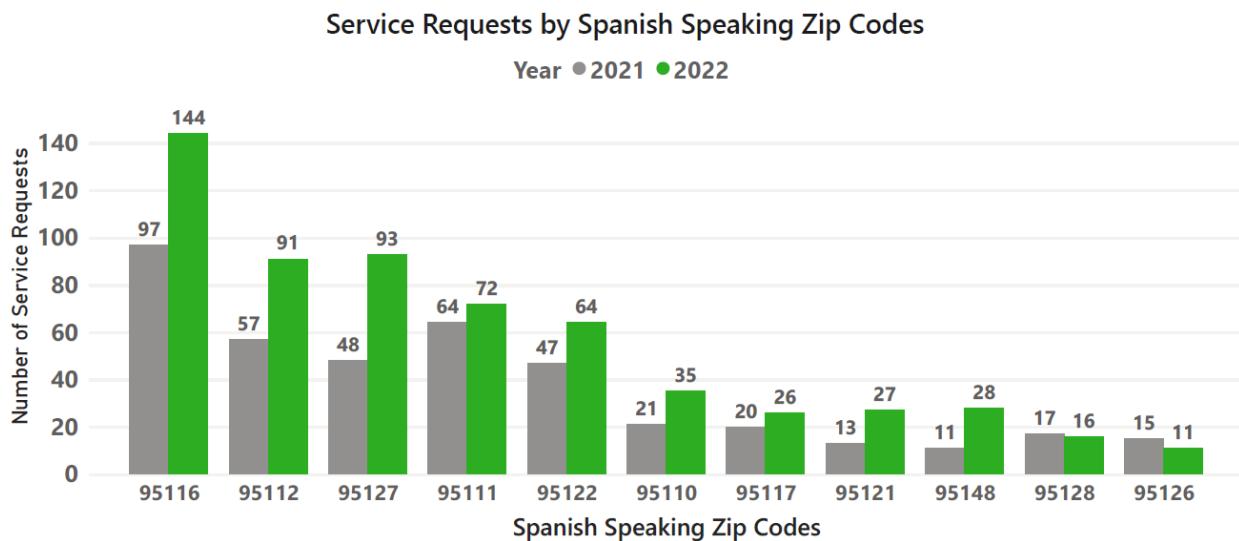

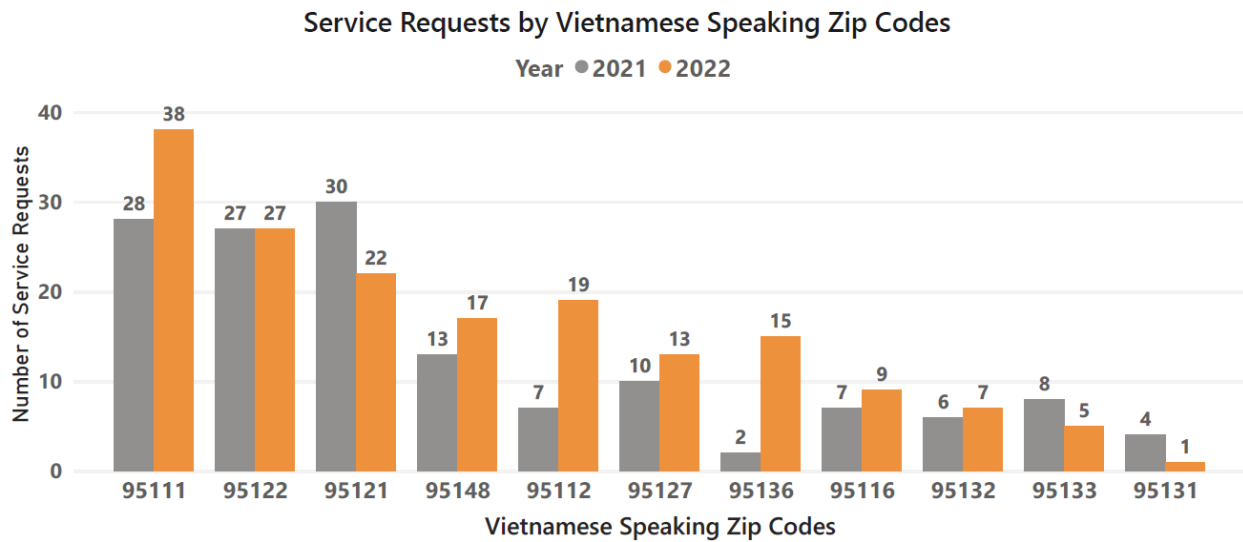**Figure 5 - 2021-22 Service request comparison before and after targeted outreach in Spanish**

Service Requests by Vietnamese Speaking Zip Codes
Year ● 2021 ● 2022

Figure 6 - 2021-22 Service request comparison before and after targeted outreach in Vietnamese

Conclusion and Recommendations

Increasing language access to City services is a multistep approach that requires strategic planning. In the case of the SJ311 platform, this included experimenting with different tools and using data-driven outreach strategies. Translation was not a simple process and posed many unforeseeable challenges, which required sophisticated solutions and human-intervention. Once quality Spanish and Vietnamese translation was available, outreach was necessary to increase access to service to non-English speaking communities. A mix of low-cost and high-cost methods were used to increase awareness of SJ311's services to Spanish and Vietnamese speaking communities. As a result of continuous outreach, an overall increase in service request/users between the 2021 and 2022 calendar years occurred.

Based on this case study of integrating Vietnamese and Spanish, recommendations for expanding to new languages on SJ311 are as follows:

- Allot additional time and resources to implementing any Asian languages – including creating a robust training/test data and rigorous evaluation criteria
- Create ways to track effectiveness of non-digital outreach strategies

As SJ311 is expanding its services, language access becomes increasingly important to ensure residents with the greatest needs are prioritized.