

Relationship Between Seattle Pet Licensing & Income by Zip Code

ETL Report

Team Members:

1. Oleksandra Ianchevska
2. Surabhi Sood
3. Zina Zachmann

Extract			
File Name	Source	Alternative Source	Format
seattle_pet_licenses.csv	https://www.kaggle.com/aaronschlegel/seattle-pet-licenses	https://data.seattle.gov/Community/Seattle-Pet-Licenses/jguv-t9rb	CSV
wa_incomes_zip_code.csv	https://www.kaggle.com/aaronschlegel/seattle-pet-licenses?select=wa_incomes_zip_code.csv	https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi	CSV
zip_code.csv	https://www.zipcodestogo.com/Washington/	---	CSV

Transform

File Name	Data Frame Name	Renamed Columns	Col. Data Type Change	Rejected Data	Deleted Column	Merged	Aggregation
seattle_pet_licenses.csv	pet_licenses_data	'zip_code' → 'Zip Code'	<u>Change Column</u> "Zip Code" = string	<u>Drop</u> incorrect zip codes	*secondary_breed * animal_s_name	zip_code_data, pet_licenses_data, left_on="Zip Code", right_on="Zip Code"	None
wa_incomes_zip_code.csv	income_data	None	<u>Change Column</u> "Zip Code" = string	None	*Number of single returns *Number of single returns *Number of joint returns *Number of head of household returns *Number with paid preparer's signature *Number of exemptions *Number of dependents *Total prepared returns	target_data, income_data, left_on="Zip Code", right_on="Zip Code"	None

					*Number of volunteer income tax assistance (VITA) prepared returns *Number of tax counseling for the elderly (TCE) prepared returns *Number of volunteer prepared *Number of refund anticipation loan returns *Number of refund anticipation check returns *Number of elderly returns [3] *Adjusted gross income (AGI) [4] *Number of returns.1		
zip_code.csv	zip_code_data	None	None	None	None	zip_code_data, pet_licenses_data, left_on="Zip Code", right_on="Zip Code"	None

N/A	target_data	None	<u>Change Column</u> 'license_number' = integer <u>Parsing</u> “License Issue Date” column into year only	<u>Drop</u> NaN <u>Drop</u> duplicates	*Adjusted Gross Income	target_data, income_data, left_on="Zip Code", right_on="Zip Code"	<u>Columns:</u> "Zip Code" "License Issue Date" "License Number" "Primary Breed" "Species" <u>Sum Columns:</u> “Number of returns” “Total Income Amount”
-----	-------------	------	---	---	---------------------------	--	---

* denotes a new column

Load		
Final Data Base	Table	Data Frame
PostgreSQL using SQLAlchemy	license_zip	target_data
HEROKU	daoqldnihoiui	target_data

The data has been prepared for analysis of correlation between a number of pet licenses issued and income by a zip code in the city of Seattle.

