



Data gathering

1. Download files from US census data as CSV
2. Upload files into Azure blob containers
3. Use an Azure data factory pipeline to run azure databricks
4. Scrape data from internet in databricks
5. Import CSV files from blob containers
6. Run Kafka producers and consumers in databricks
7. Output data from any imported, scraped or consumed to azure data lake and azure SQL databases
8. Continue data factory pipeline and copy data from data lake to sql database
9. Import data from databases in PowerBI and apply machine learning with PowerBI Python

rates	
	Age int
	Gender nvarchar(20)
	\$250,000 float
	\$500,000 float
	\$750,000 float
	\$1,000,000 float

life_insured	
	year int
	percentage int

ownership_gap	
	year int
	percentage int

why_insured	
	Reason to hold life insurance nvarchar(100)
	2017 int
	2019 int
	change int negative

why_not_insured	
	Reason to hold life insurance nvarchar(100)
	percentage int

iii	
	Payout Category nvarchar(100)
	2016 int
	2017 int
	2018 int
	2019 int
	2020 int