OCCUPATION DATA

Data Science Project

Dustin Fraser

AGENDA

1. Overview
2. Data Ethics
3. Business Problem
4. Performance Measure
5. Getting the Data
6. Exploring the Data for Insights
7. Exploring the Models
8. Fine Tuning
9. Presenting the Solution

OVERVIEW OF BUSINESS OBJECTIVE

- What is the estimated number of employed and unemployed individuals in Region One of a small nation state?
- What are the specific occupation types to support government policies such as job assistance programs?
- What is the estimated unemployment rate as an overall indicator of the health and wealth of the region?

DATA ETHICS

- The data was obtained from the country's electoral commission.
- The commission published basic demographic data on individuals of voting age (name, address, occupation, and a unique identifier).
- The variables of interest were the occupation and identifies.
- For this project, the other variables were masked to protect the individuals although this information is publicly available.

BUSINESS PROBLEM

- What is the number of unemployed individuals in Region One?
- What are the specific occupation types to support government policies such as job assistance programs?
- What is the unemployment rate as an overall indicator of the health and wealth of the region?

ANALYTIC CONSIDERATIONS

Analytic Type: Descriptive – use the existing snapshot of the data to describe the current state of the region. May be used in a later model (neural network) and to predict changes.

Current State: No such program exist in the country.

Learning Approach: Semi-Supervised (Clustering & Simple Linear Regression).

Data Processing Style: Batch – reading data from *.csv.

PERFORMANCE MEASURE

Key Metrics:

- Examine the number of employed vs unemployed.
- Reduction in the number of unemployed individuals.
- Number of unemployed will be tracked and benchmarked against government programs over quarterly periods.
- Monitor changes to currently employed numbers.

DATA COMPARISON

Comparison Method: Euclidean Distance & Silhouette

Expertise: No expertise but capacity development is required.

Manual Solution: Individuals in need may randomly walk into financial assistance department. Financial help data is managed in localized records and based on a speculative budget.

PERFORMANCE MEASURE - ASSUMPTIONS

Key Metrics:

- Employment status is primarily based on investment in farming and mining.
- Climate, transportation, business conditions affect the industries.
- The most important datapoint is unemployment but later work may include “Housewife” to determine household financial disposition.
- The unemployment measure indicates social and economic risks of the region.
- Cost to government.

EXPLORING THE DATA FOR INSIGHTS - DICTIONARY

Fig 4. Data of interest extracted, converted into a dictionary, and other variables dropped

```
Counter({'-': 7980, 'FARMER': 3021, 'MINER': 1825, 'HOUSEWIFE': 912, 'LABOURER': 434, 'TEACHER': 391, 'HOUSEKEEPER': 317, 'CARP
ENTER': 242, 'BUSINESSMAN': 194, 'FISHERMAN': 184, 'COOK': 132, 'DRIVER': 132, 'OPERATOR': 124, 'SECURITY GUARD': 101, 'MECHANI
C': 96, 'BUSINESSWOMAN': 89, 'SAW OPERATOR': 67, 'MASON': 57, 'LOGGER': 48, 'SALES CLERK': 47, 'SHOPKEEPER': 46, 'HEALTH WORKE
R': 41, 'CHAINSAW OPER.': 37, 'HOUSE KEEPER': 37, 'PORTER': 36, 'DOMESTIC': 35, 'CLEANER': 34, 'VENDOR': 33, 'SURVEYOR': 27, 'S
EAMSTRESS': 27, 'WELDER': 26, 'SALESMAN': 24, 'NURSE': 21, 'CHAIN SAW OPER.': 20, 'HOUSE-KEEPER': 18, 'CLERK': 18, 'PORK KNOCKE
R': 18, 'JOINER': 17, 'SHOP KEEPER': 16, 'BOAT CAPTAIN': 15, 'DOMESTIC WORKER': 15, 'CONDUCTOR': 14, 'OUTBOARD OPER.': 14, 'POL
ICE': 13, 'MAID': 13, 'ELECTRICIAN': 13, 'SUPERVISOR': 12, 'SEAMAN': 12, 'ACCOUNTANT': 12, 'BARBER': 12, 'SALESGIRL': 12, 'ACCO
UNTS CLERK': 12, 'NONE': 11, 'TAILOR': 11, 'STEVEDORE': 11, 'CHARWOMAN': 11, 'MANAGER': 10, 'HOMEMAKER': 10, 'BOATMAN': 10, 'SA
WMAN': 10, 'SALESWOMAN': 9, 'LINESMAN': 9, 'CHEF': 8, 'WARD MAID': 8, 'SECURITY': 8, 'POLICEMAN': 8, 'PAINTER': 8, 'CASHIER':
8, 'RECEPTIONIST': 8, 'STOREKEEPER': 8, 'CONSERVATIONIST': 8, 'MEDEX': 7, 'HAIRDRESSER': 7, 'MOTOR MECHANIC': 7, 'SALES GIRL':
7, 'TRADER': 7, 'GEO TECHNICIAN': 7, 'MICROSCOPIST': 7, 'BUS CONDUCTOR': 6, 'SAILOR': 6, 'SOLDIER': 6, 'WAITRESS': 6, 'STORE CL
ERK': 6, 'COAST GUARD': 6, 'GUARD': 6, 'TYPIST': 6, 'BOAT OPERATOR': 6, 'CHAIN SAW OPER': 6, 'ORE SAMPLER': 6, 'ADMINISTRATOR':
5, 'BUS DRIVER': 5, 'HANDYMAN': 5, 'LOGGING CLERK': 5, 'DOMESTIC MAID': 5, 'PASTOR': 5, 'SECRETARY': 5, 'TECHNICIAN': 5, 'MISSI
ONARY': 5, 'CRAFTMAN': 5, 'MACHANIC': 5, 'BARTENDER': 5, 'LINE CUTTER': 5, 'GOLD MINER': 4, 'TOSHAO': 4, 'CRAFTSMAN': 4, 'SALES
MANAGER': 4, 'COSMETOLOGIST': 4, 'BAKER': 4, 'GOLDSMITH': 4, 'DENTIST EXT.': 4, 'DISEASE CON OFF': 4, 'NURSE ASSISTANT': 4, 'CH
AINSAW OPER': 4, 'STORES CLERK': 4, 'ASST. REG. OFF.': 4, 'DRILLER HELPER': 3, 'PUMP OPERATOR': 3, 'CATERER': 3, 'FORESTER': 3,
'SALESCLERK': 3, 'COORDINATOR YCG': 3, 'CHAFFEUR': 3, 'PLUMBER': 3, 'ADMIN. ASST.': 3, 'POSTAL AGENT': 3, 'OFFICE ASST': 3, 'SE
LF EMPLOYED': 3, 'WARD ORDERLY': 3, 'CARE TAKER': 3, 'SOCIAL WORKER': 3, 'STENOGRAPHER': 3, 'GARDENER': 3, 'Occupation': 3, 'SH
OP ASSISTANT': 3, 'TECHNOLOGIST': 3, 'JANITOR': 3, 'C HEALTH WORKER': 3, 'FIELD ASST MANG': 3, 'ORE. SAMPLER': 3, 'GEO. TECH.':
3, 'VET OFFICER': 3, 'OUTBOARD OPER': 3, 'CHIEF COOK': 3, 'EXCAVATOR OPER.': 3, 'MECHANIC DRIVER': 2, 'FLIGHT AGENT': 2, 'DENTI
ST': 2, 'BUSINESS MAN': 2, 'HOUSEMAKER': 2, 'BUSINESS WOMAN': 2, 'VAN DRIVER': 2, 'PROSPECTOR': 2, 'SECURITY SUP': 2, 'DENTAL T
ECH.': 2, 'GEOLOGIST': 2, 'TOLL CLERK': 2, 'CANE CUTTER': 2, 'SHOP-KEEPER': 2, 'ENGINEER': 2, 'CONFECTIONER': 2, 'FITTER': 2,
'MIN OF RELIGION': 2, 'REGIST. CLERK': 2, 'LOGGING': 2, 'FOGGMAN': 2, 'VILLAGE CAPTAIN': 2, 'FRAMER': 2, 'FOREMAN': 2, 'SUPERIN
TENDENT': 2, 'CHAINSAW OPERA.': 2, 'REG. CLERK': 2, 'LAUNDRESS': 2, 'CAPTAIN': 2, 'LAND ADMIN. OFF': 2, 'CHAUFFEUR': 2, 'CLERK
OFFICER 2': 2, 'TYPIST CLERK': 2, 'WARDMAID': 2, 'COUNTER CLERK': 2, 'DOCTOR': 2, 'HYDROMET TECH.': 2, 'SHOP PROPRIETOR': 2,
'C.S.R': 2, 'DESIGNER': 2, 'POST CLERK': 2, 'COORDINATOR': 2, 'REG LEADER': 2, 'WATCHMAN': 2, 'REG. CLERK ONE': 2, 'POSTAL APP
R.': 2, 'MIDWIFE': 2, '--': 2, 'CABBAGE CUTTER': 2, 'MINE SURVEYOR': 2, 'CHAIN SAW OP': 2, 'STUDENT NURSE': 2, 'MEDICAL TECH.':
2, 'C.H.WORKER': 2, 'CAPTAIN ENG.': 2, 'VENEER': 2, 'CUSTOMER REP.': 1, 'COUNSELLOR': 1, 'POLICEWOMAN': 1, 'MANAGER CWC': 1, 'P
OST APPRENTICE': 1, 'COMM H/TH WRK.': 1, 'COMM SUPP OFF': 1, 'TRIBUTOR': 1, 'ASSISTANT COOK': 1, 'GEOLOGY TECH.': 1, 'SMALL WOR
KER': 1, 'SALES REP': 1, 'MULTI PURP TECH': 1, 'GOLD SMITH': 1, 'A.R.O': 1, 'DISC JOCKEY': 1, 'BUTCHER': 1, 'SALESBOY': 1, 'PLA
NT OPERATOR': 1, 'HOSP. ATTENDANT': 1, 'COM. HEALTH WKR': 1, 'CONTRACTOR': 1, 'ENVIRON OFFICER': 1, 'DOTS WORKER': 1, 'CARTMA
N': 1, 'PSYCHIATRIC ASS': 1, 'C.H.W': 1, 'LIVESTOCK OFF.': 1, 'UTILITY OPT.': 1, 'LINES MAN': 1, 'TOSHAU': 1, 'POLICE OFFICER':
1, 'C/HEALTH WORKER': 1, 'VICE CAPTAIN': 1, 'CHAINSAW OPR.': 1, 'TRACT OPERATOR': 1, 'CHAINSAW OPT.': 1, 'TR. OPERATOR': 1, 'RE
G. CHAIRMAN': 1, 'CLERICAL OFFICE': 1, 'FLIGHT ATT.': 1, 'DRIVER/MECHANIC': 1, 'SHOP ATTENDANT': 1, 'SALESPERSON': 1, 'CARE GIV
ER': 1, 'BOAT HAND': 1, 'OVERSEER': 1, 'SECURITY SUP.': 1, 'HOUSE-WIFE': 1, 'REGIST. OFFICER': 1, 'PREACHER': 1, 'MEDICAL DOCTO
R': 1, 'CIVIL ENGINEER': 1, 'MIN.OF RELIGION': 1, 'POSTAL CLERK': 1, 'CLERK 3': 1, 'NURSE MANAGER': 1, 'PRIEST': 1, 'CLERICAL O
FF.': 1, 'FIELD ASSISTANT': 1, 'ATTENDANT': 1, 'PLANT ATTENDANT': 1, 'BOAT BUILDER': 1, 'AGRICULTURIST': 1, 'FOREST OFFICER':
1, 'CONDUCTRESS': 1, 'WARDEN': 1, 'ARTIST': 1, 'LUMBERMAN': 1, 'SHOP OWNER': 1, 'SAMPLE TECH': 1, 'C D OFFICER': 1, 'CARETAKE
R': 1, 'TRACTOR OPER.': 1, 'TRAPPER': 1, 'REG. OFFICER': 1, 'MED TECHNICIAN': 1, 'CRAFTSWOMAN': 1, 'NURSING ASST.': 1, 'ASST.CR
EDIT OFF': 1, 'DORMS SUP.': 1, 'BOATHAND': 1, 'BOAT DRIVER': 1, 'MACHINE OPER.': 1, 'GEO TECH': 1, 'PHARMACIST': 1, 'TRACTOR DR
IVER': 1, 'PHOTOGRAPHER': 1, 'E/MENTALIST': 1, 'LUMBERJACK': 1, 'COM SUPPORT OFF': 1, 'RANGER': 1})
```

EXPLORING THE MODELS - SILHOUETTE ANALYSIS

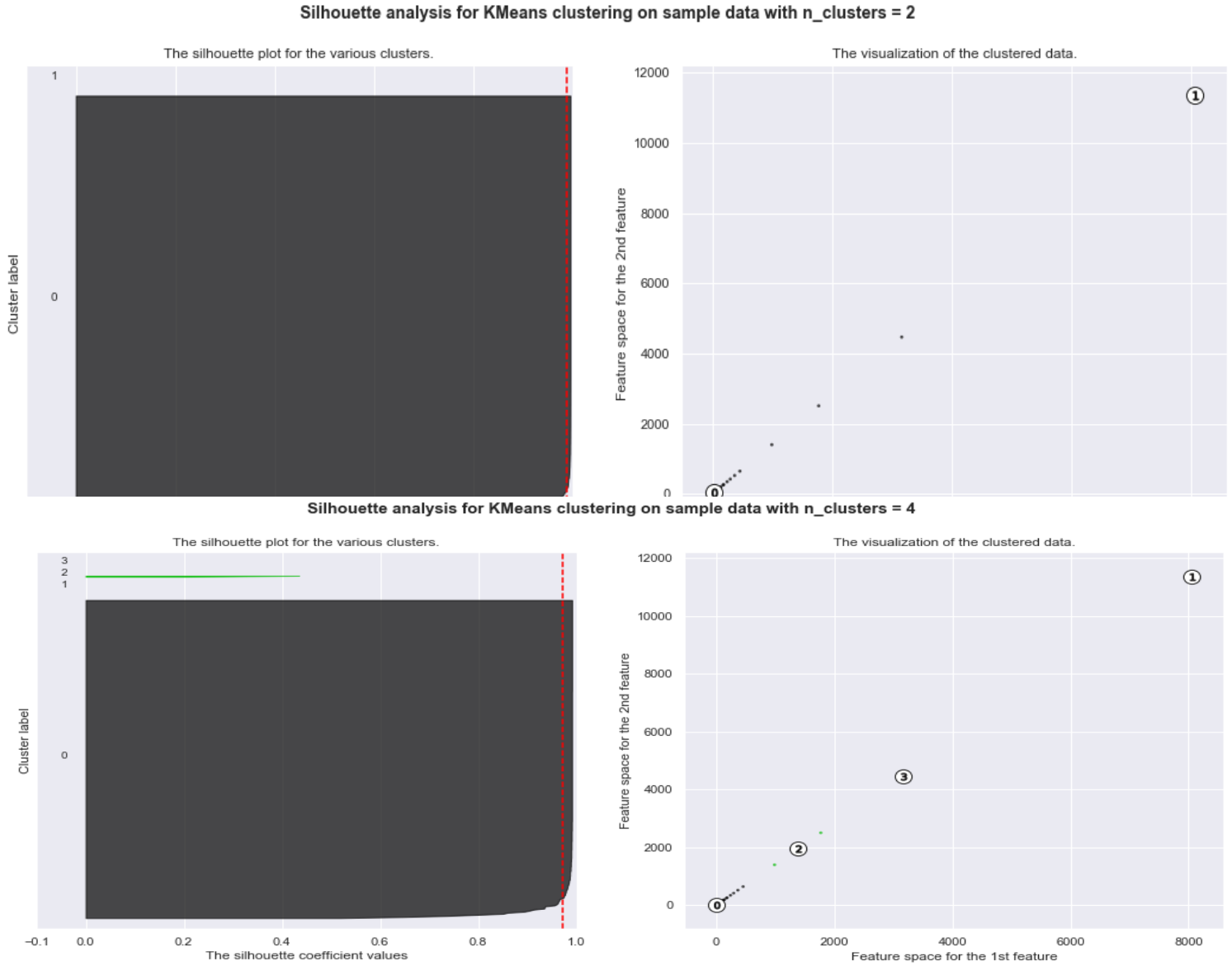

Fig 9. A study of the separation distance between the resulting clusters

THE SOLUTION - REGION ONE

- Approximately 47% of residents are unemployed.
- At this stage, the prediction model may not support the measure of joblessness but helps government policies and initiatives.
- Although the number of households were not in scope, 912 housewives are worth exploring in a qualitative study.
- The number of public servants were inadequate for the population size.
- Most of the occupations require little skills.
- Forms a great start to input into advance models to understand household income and other social and economic advancement.