PREDICT AND MAP WHEAT PRODUCTION ACROSS THE SK, BASED ON A 1938–2021 DATABASE USING AN UNSUPERVISED MACHINE LEARNING CLUSTERING APPROACH.

Viraj Hewage

INTRODUCTION

The crops, lands, and timing of 622 rural municipalities in Saskatchewan were used in a historical data analysis. Python and Geopanda used unsupervised machine learning clustering to analyze this large cohort.

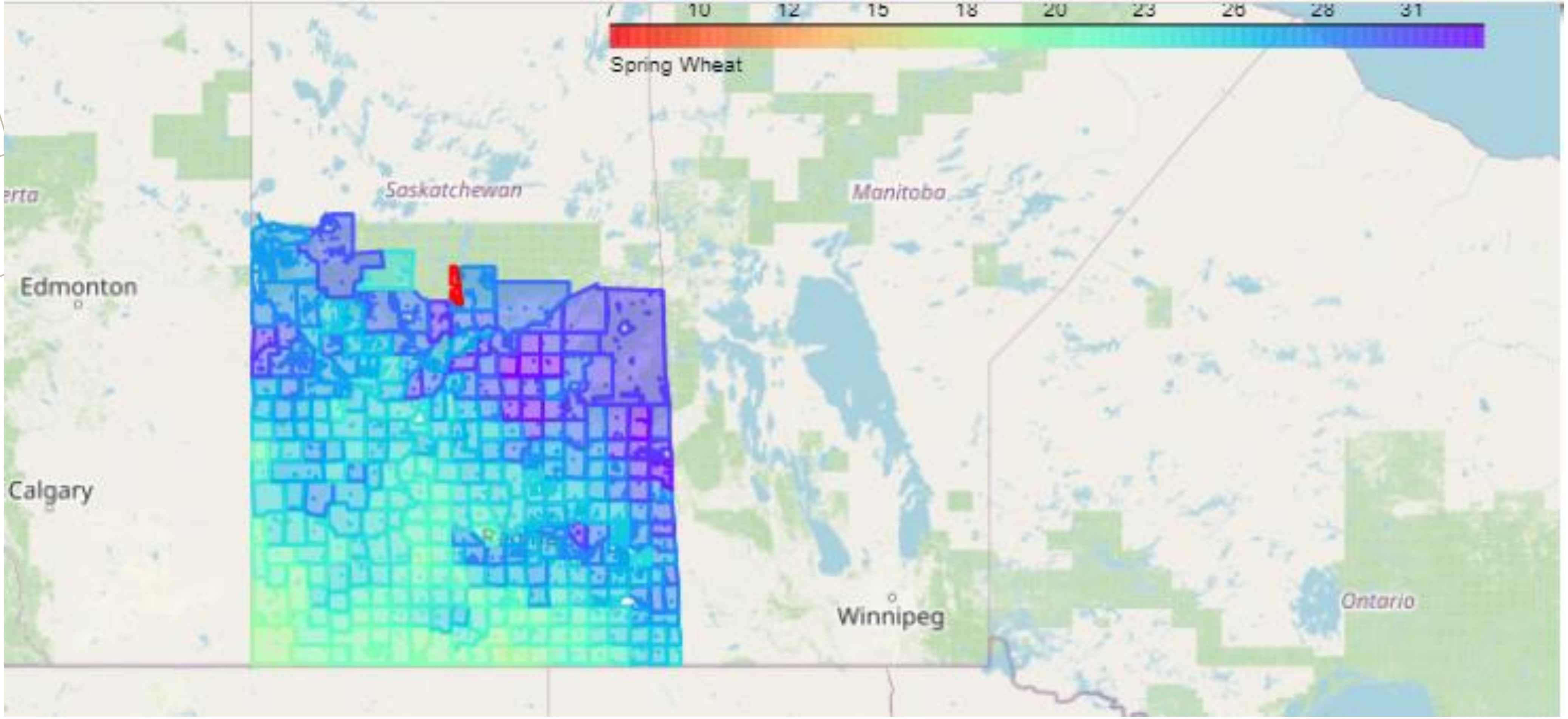
DATA BASES

rm_crop_yields_1938_2021.csv

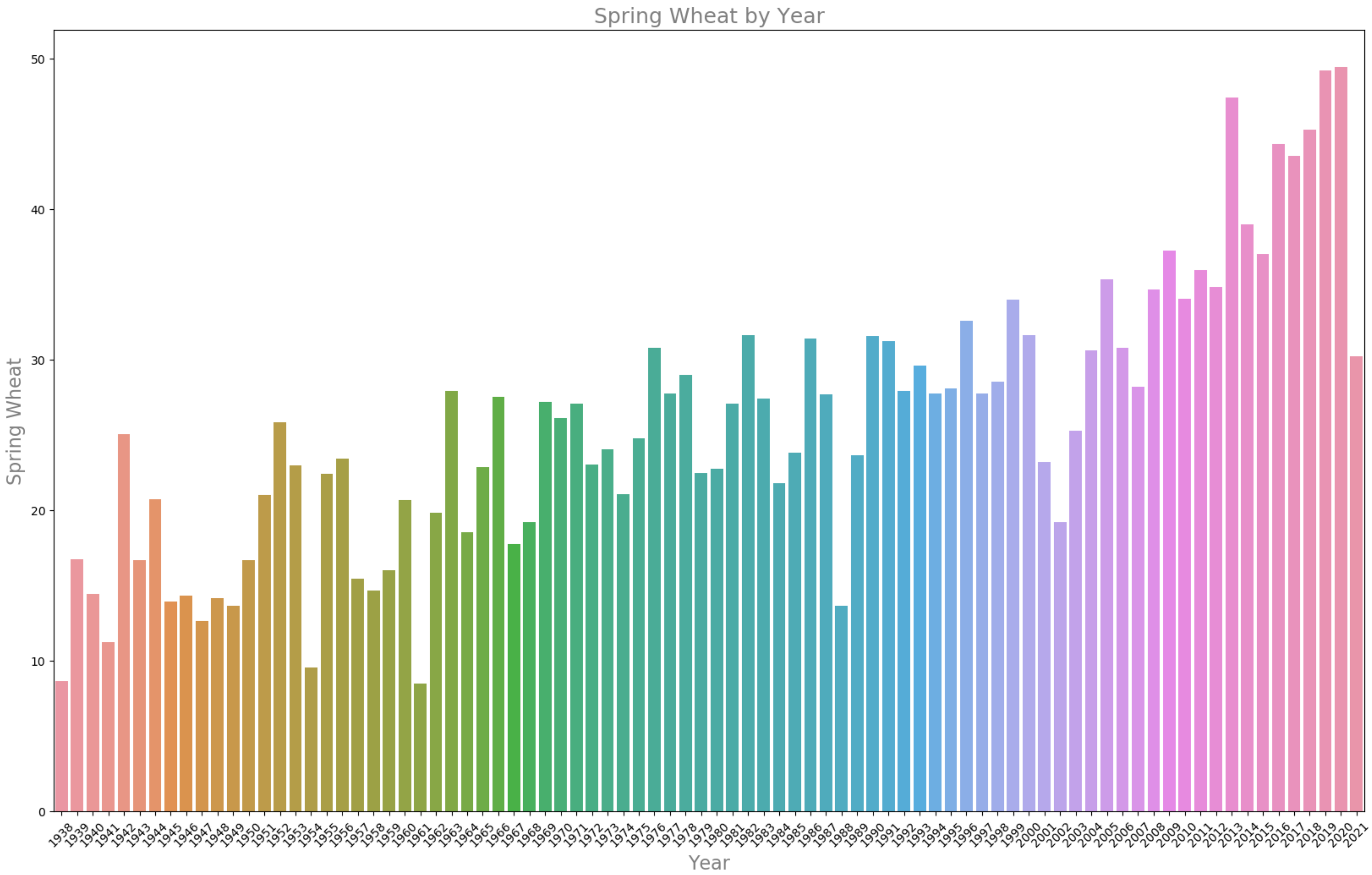
Rural Mnicipilaty.shp

DATA ANALYSIS

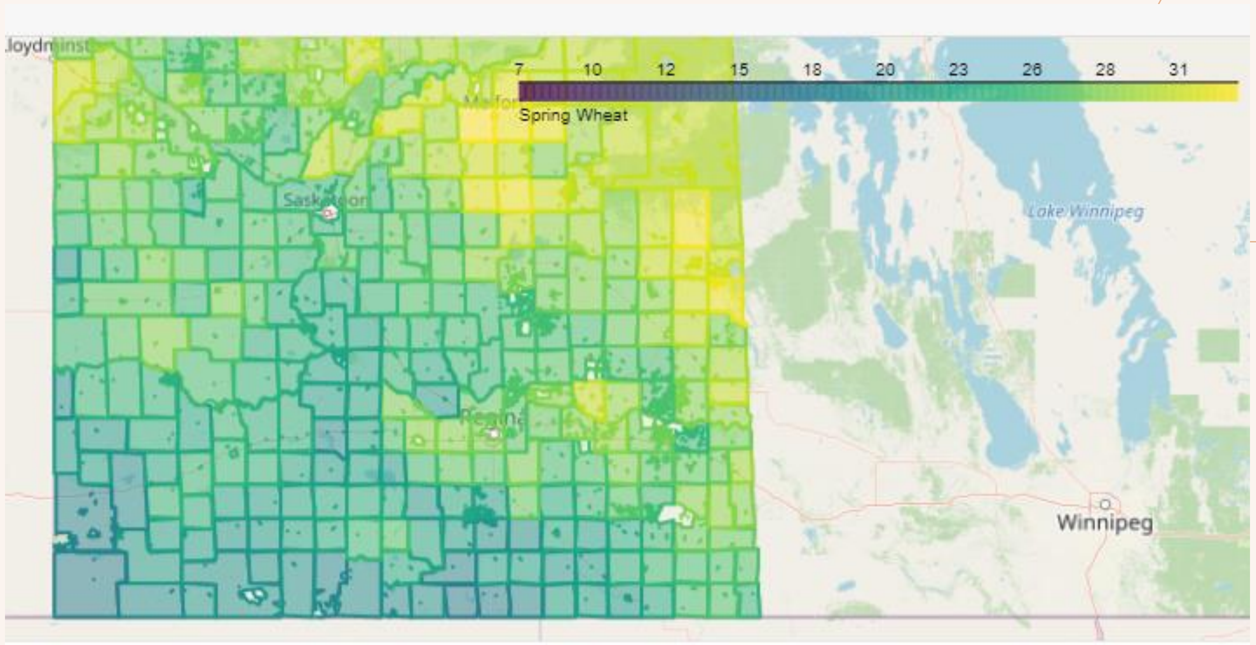
Why I chose Spring wheat data?



WHEAT PRODUCTION ACROSS THE SK, BASED ON A 1938–2021 DATABASE

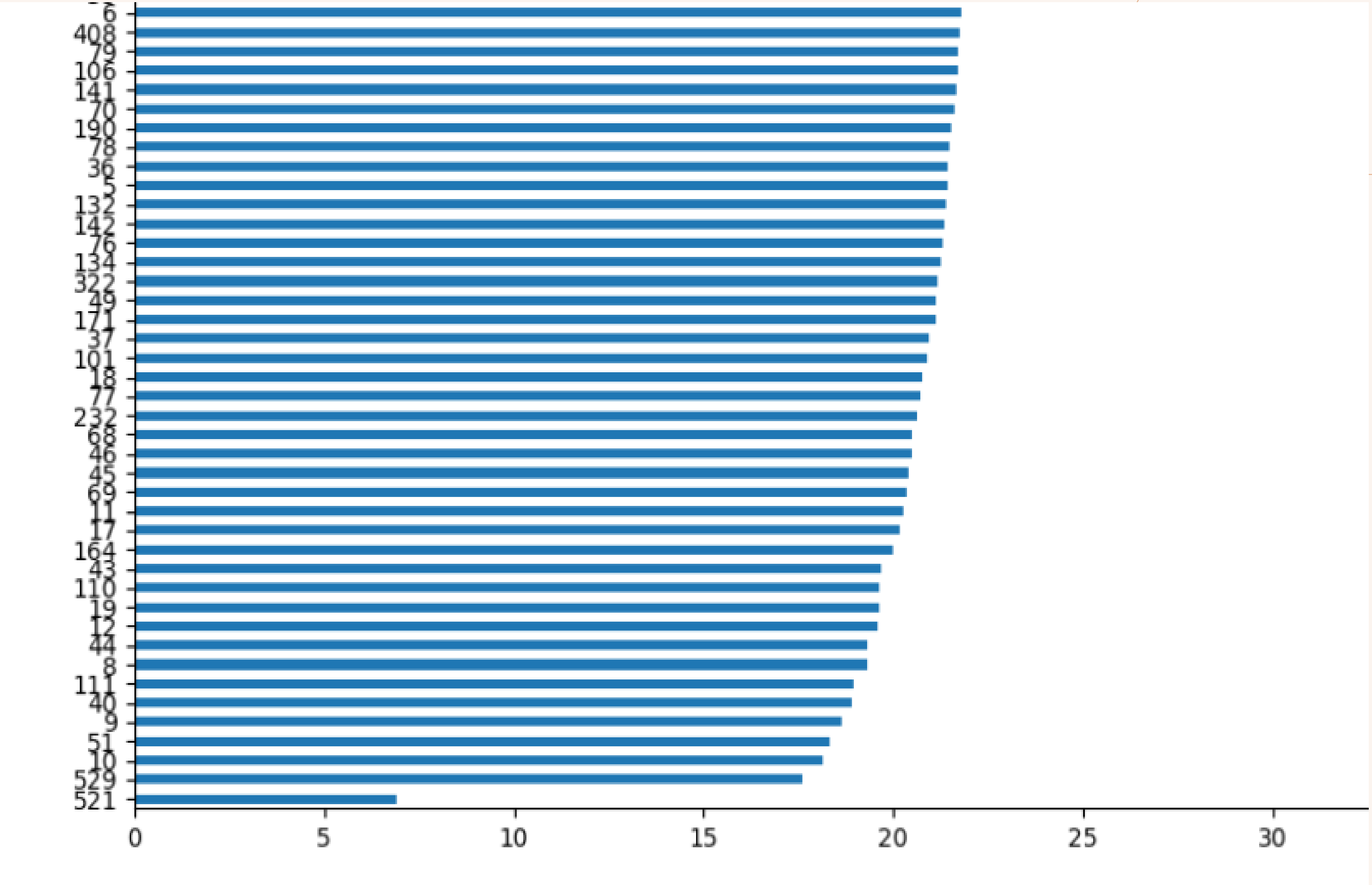


METHODOLOGY

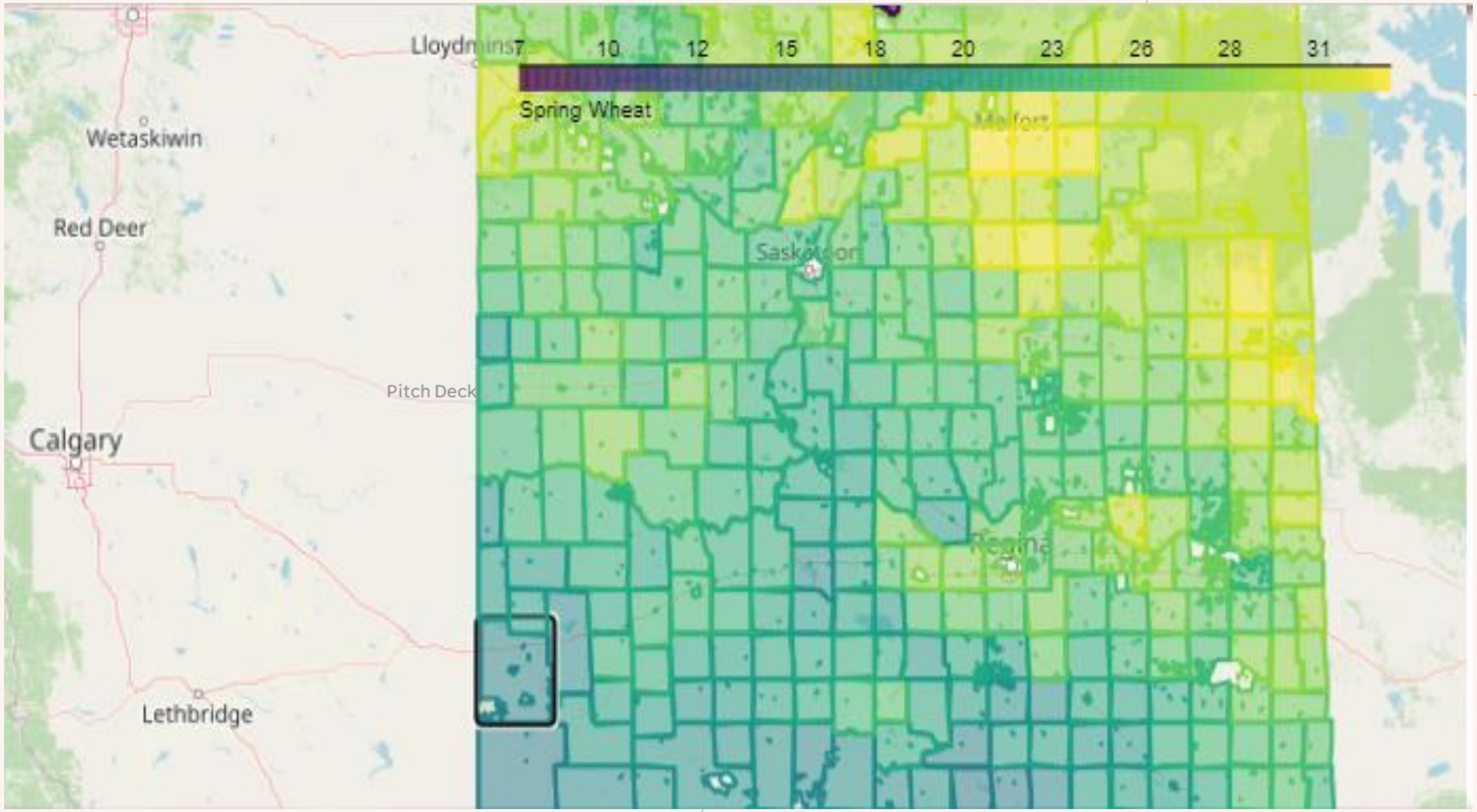


Generating Summary Statistics: Use the describe() method to generate summary statistics such as mean, standard deviation, quartiles, etc. Data were analyzed using the K-Means Clustering method as a technique for performing data groupings. Furthermore, the data classification procedure was based on the degree of each component's membership. This analysis was performed by using python (pandas, numpy, matplotlib, seaborn) and Geopandas.

FINDING A RM FOR MODELING



TIME SERIES FORECASTING RM 111 MAPLE CREEK

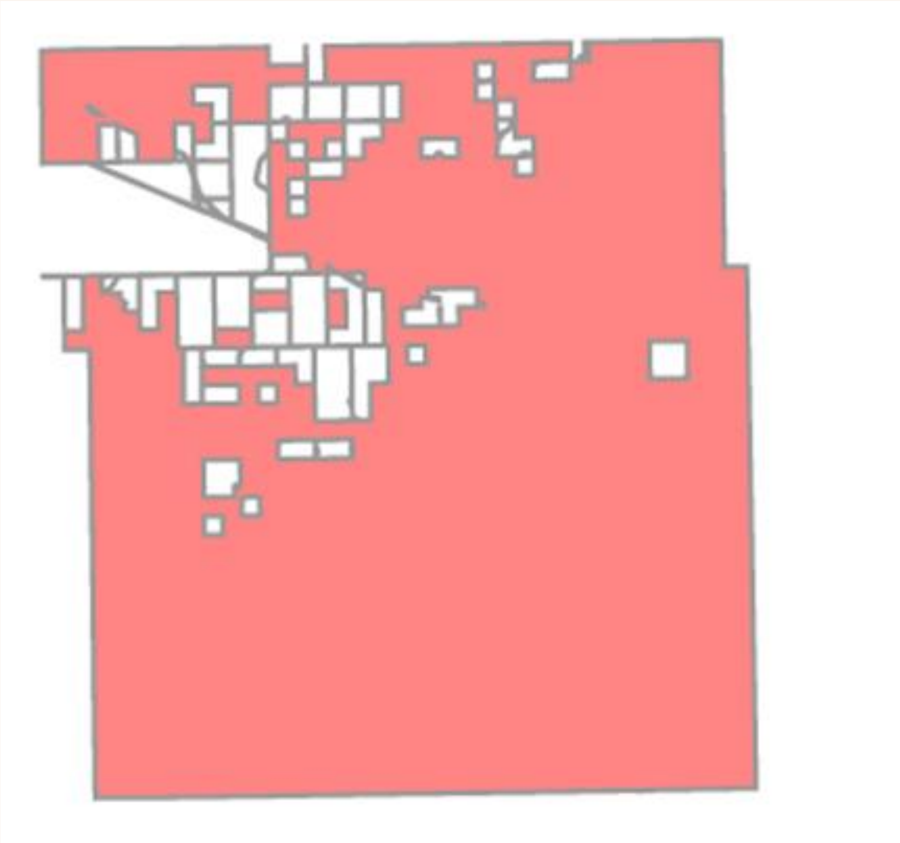


DATA MODELING FOR RM 111, MAPLE CREEK

Unsupervised machine learning algorithms are also useful for understanding the relationships between different variables in a dataset. For example, you can use clustering algorithms such as K-Means to identify clusters of similar points in a dataset

Define the features to be used for clustering

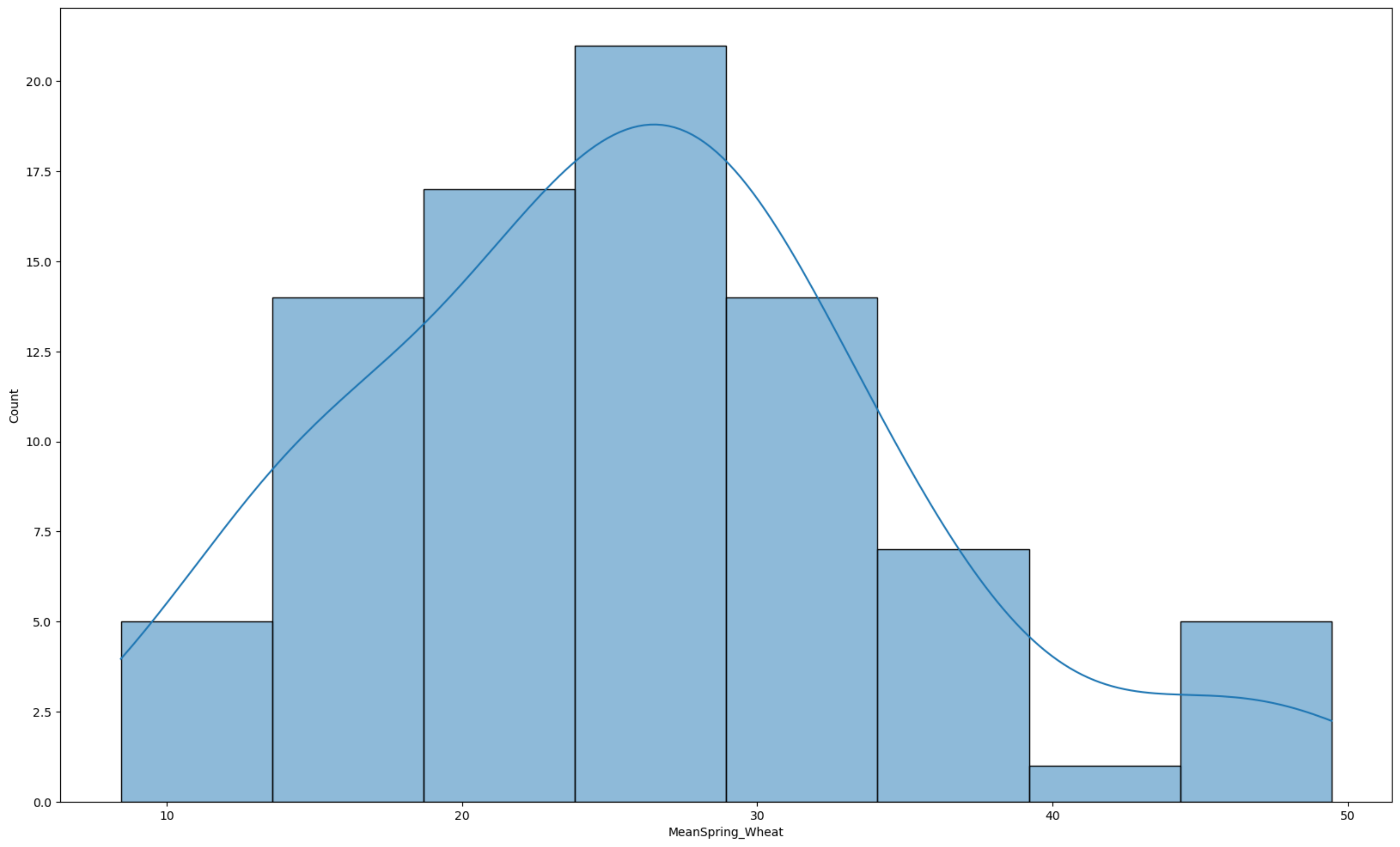
CREATE A MODEL FOR SPRING-WEAT IN RM 111 MAPLE CREEK



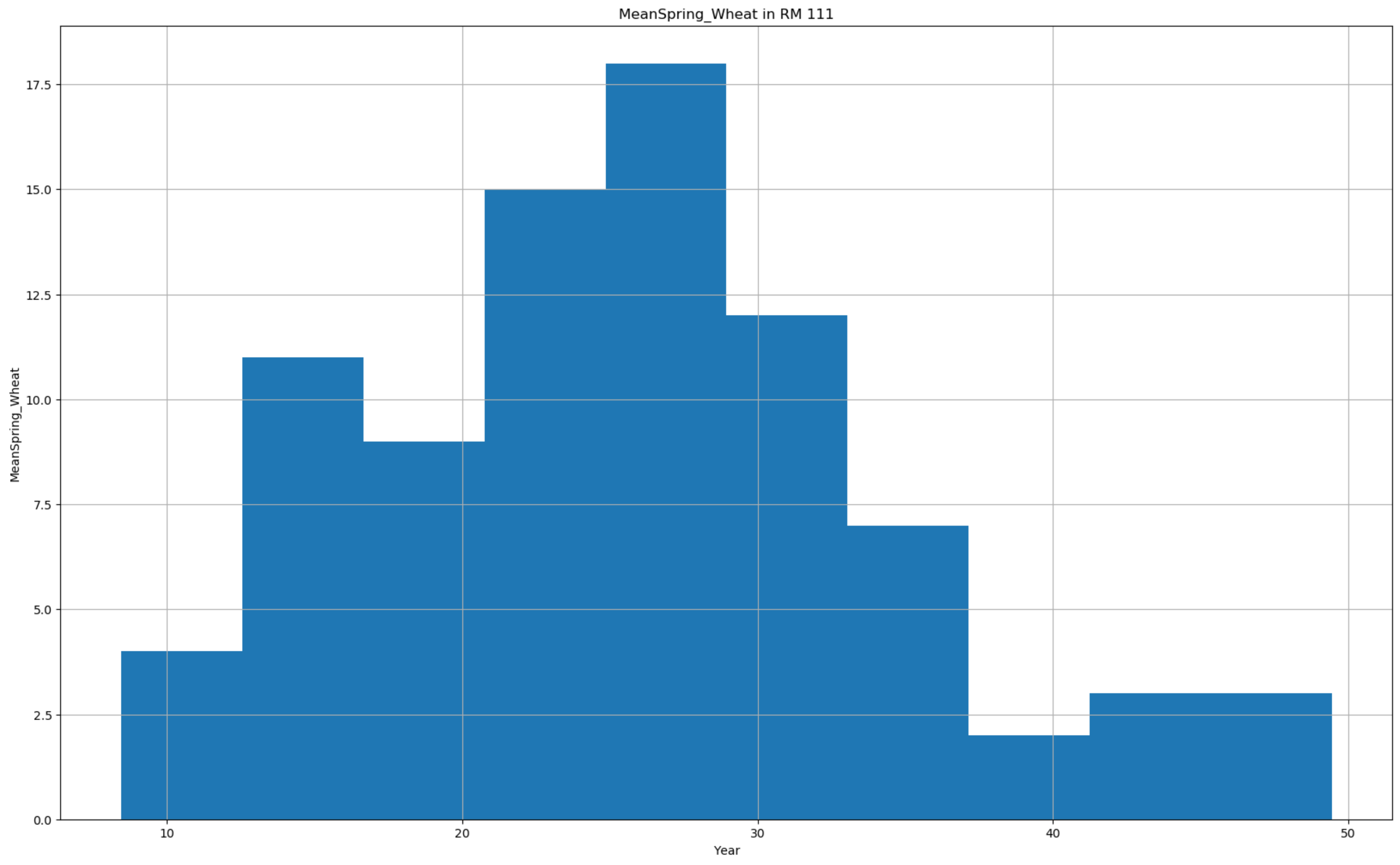
DATA MODELING FOR SK 622 RMS -HISTOGRAM

Shape (84, 6)

slope: 0.3097599939120827
intercept: -587.316443382177
r: 0.8226524204739321
p: 8.255635700786158e-22
stderr: 0.023641040235293034

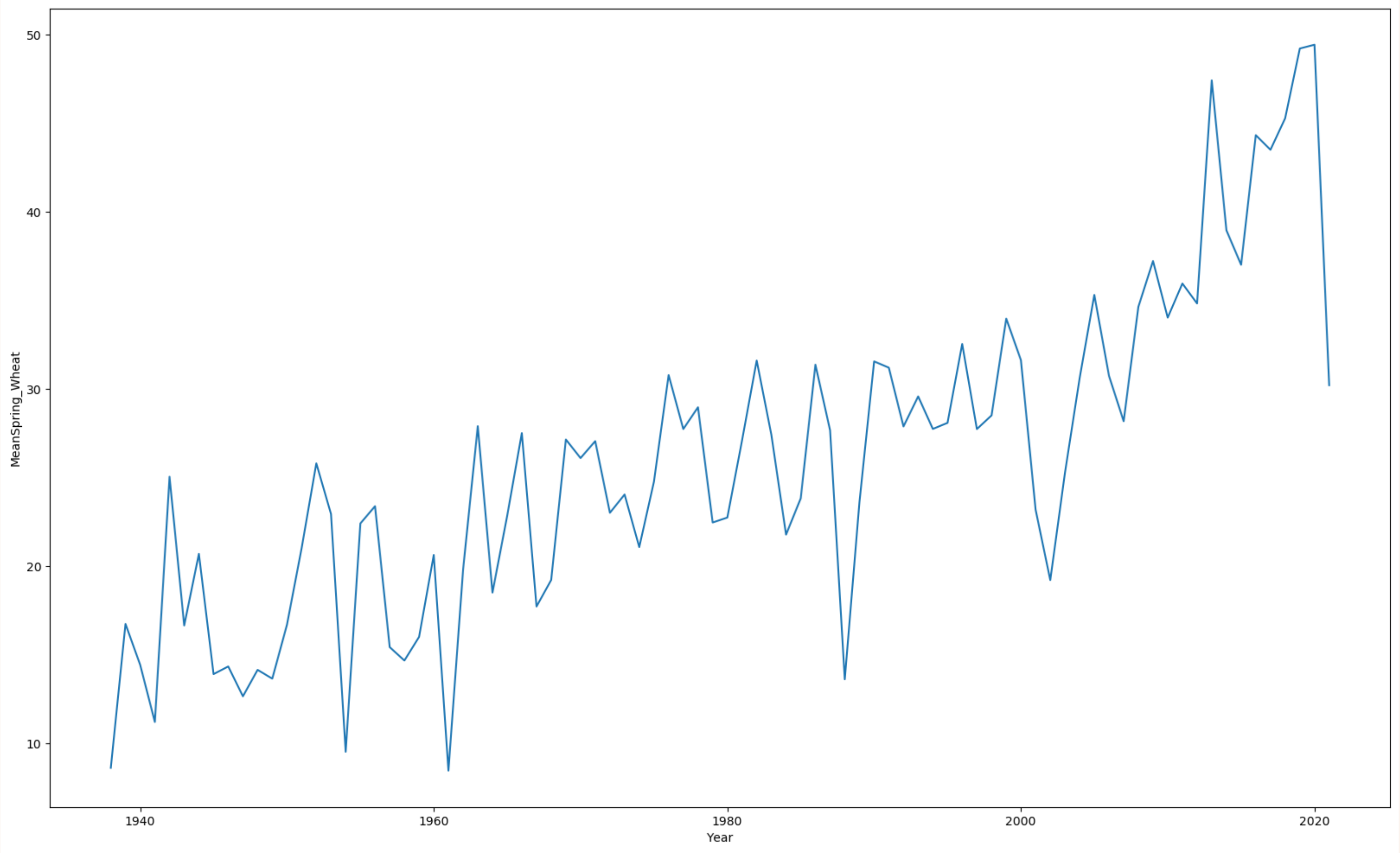


DATA MODELING FOR SK 622 RMS – BY YEAR



TIME SERIES FORECASTING RM 111 MAPLE CREEK

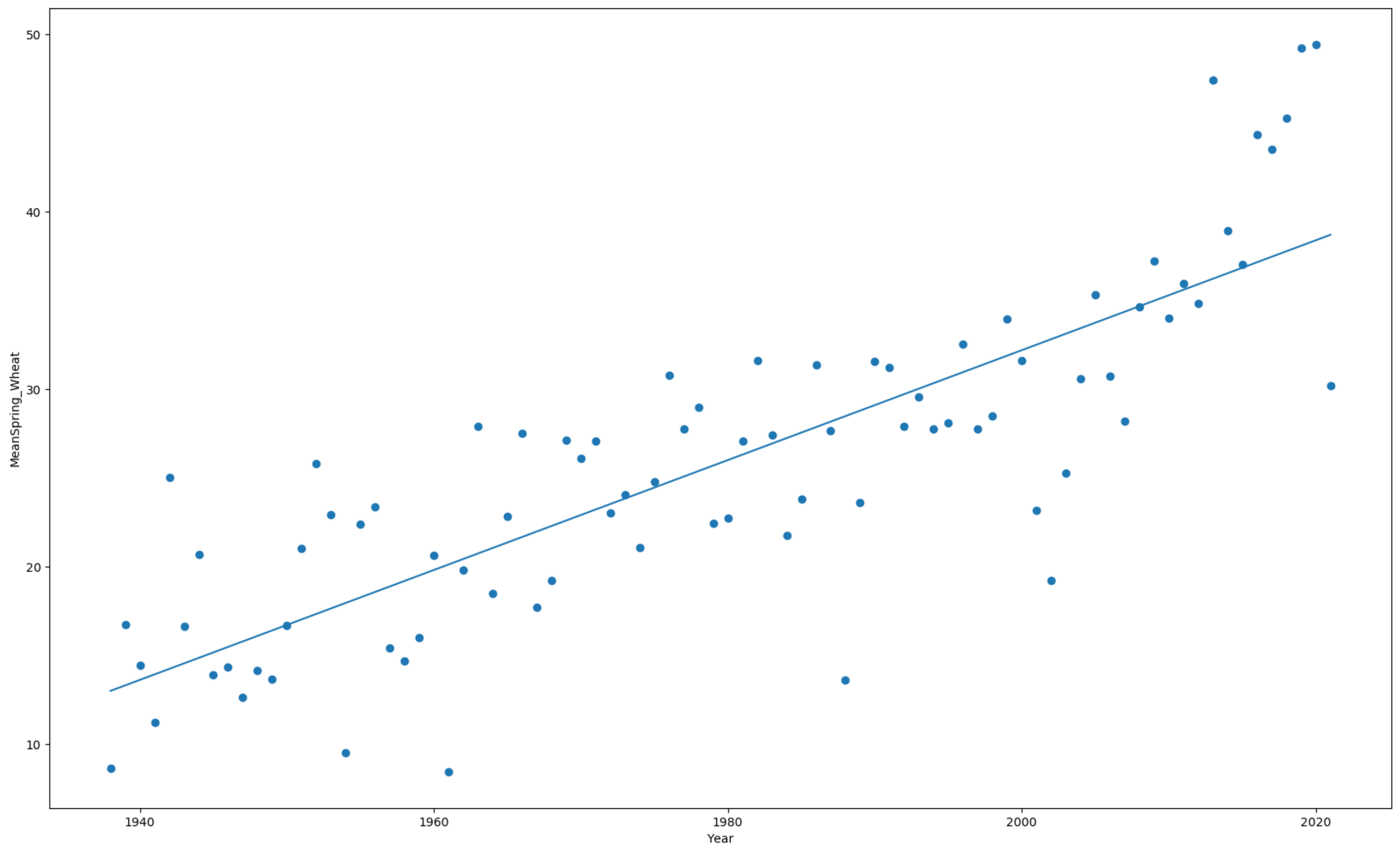
Autoregressive Integrated Moving Average (ARIMA) models facilitate demand forecasting, such as in determining future demand for spring_weats. That is because the model provides products related to time. This forecasting can also be used to predict future yield production.



LINEAR REGRESSION FORECASTING FOR SPRING_WEAT IN RM 111, MAPLE CREEK

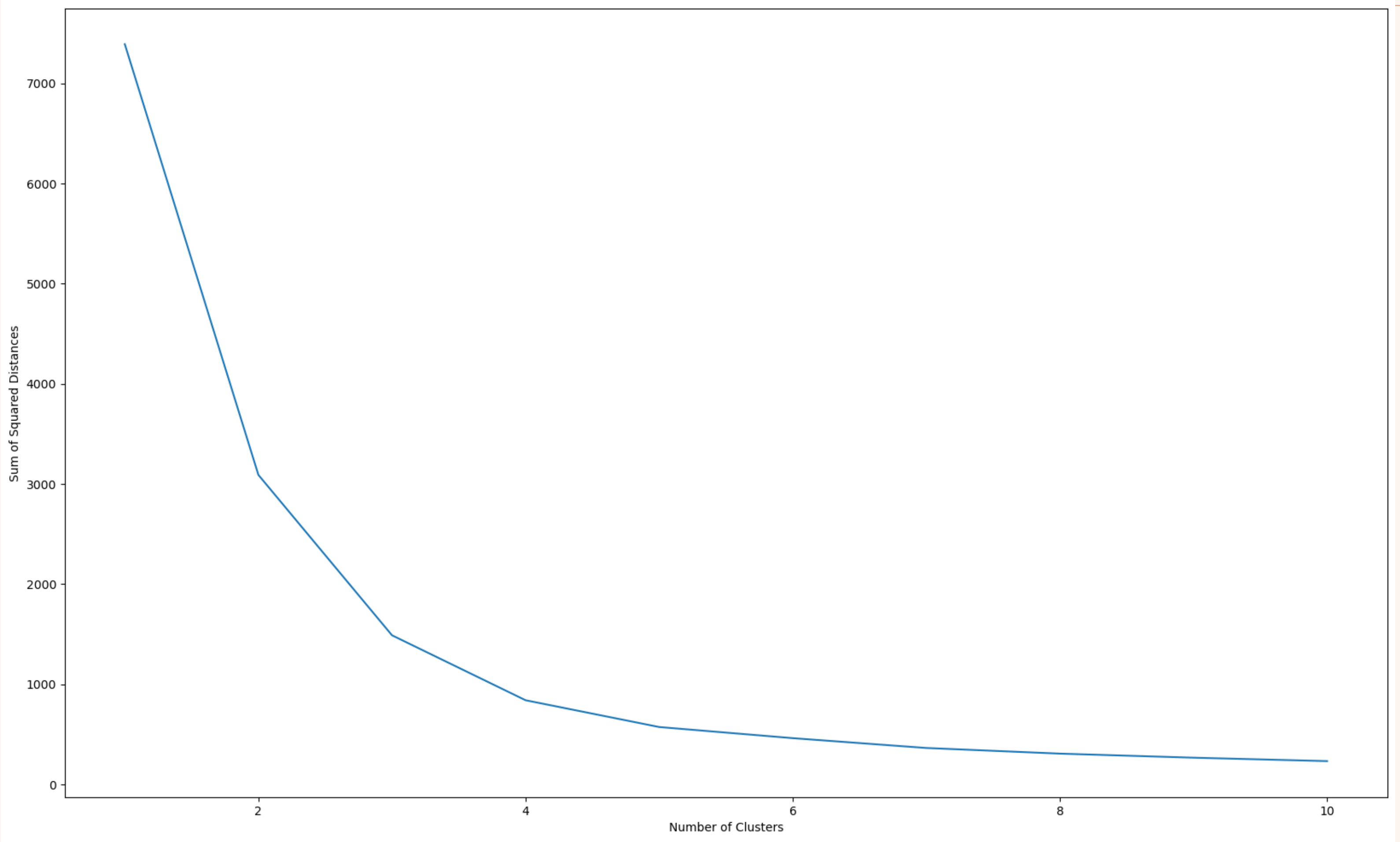
We can see that there is a positive linear relationship between amount of mean spring_weat production and time.

With this model, we can predict production increases with the time.

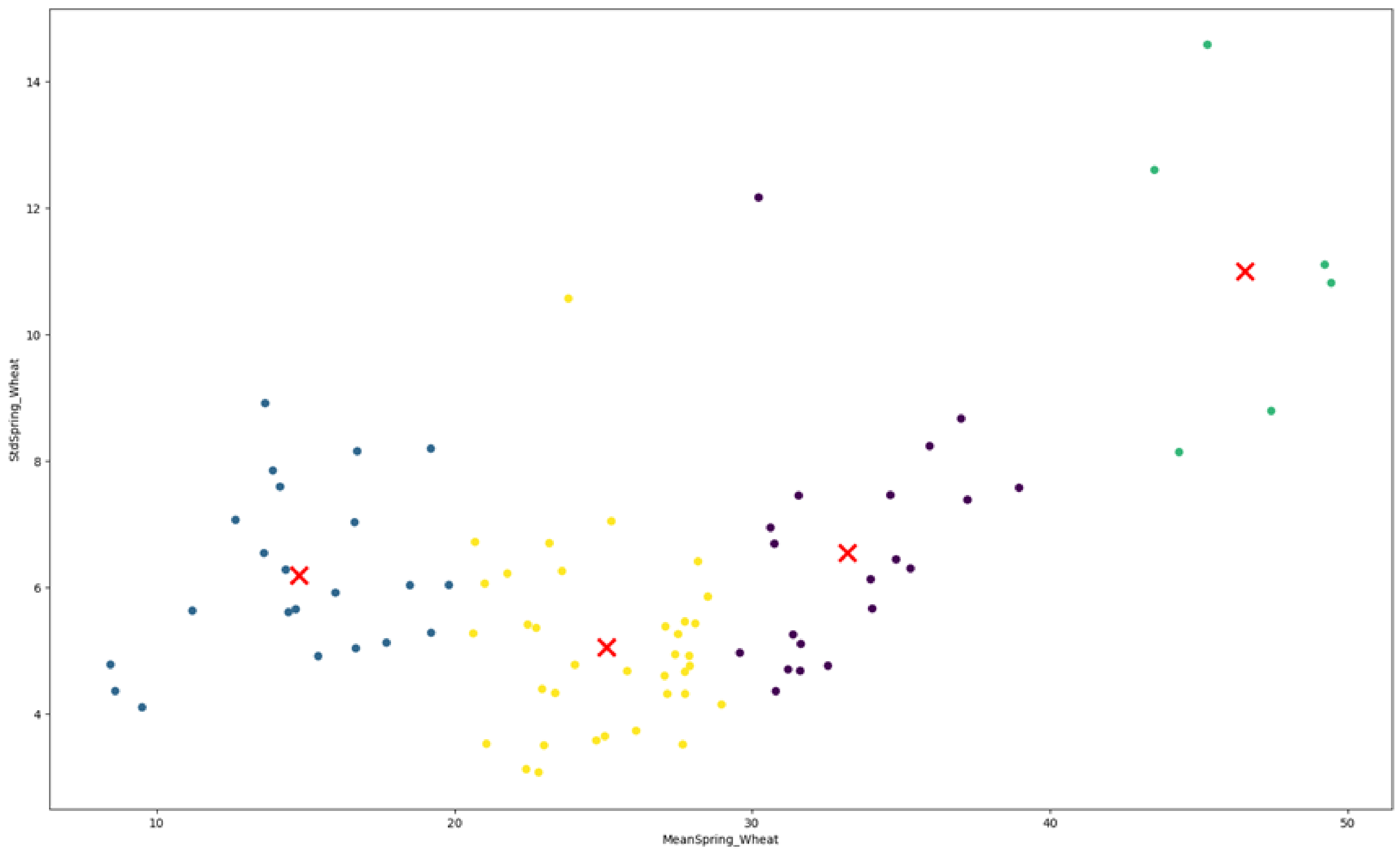


K-MEANS CLUSTERING WITH SCIKIT-LEARN

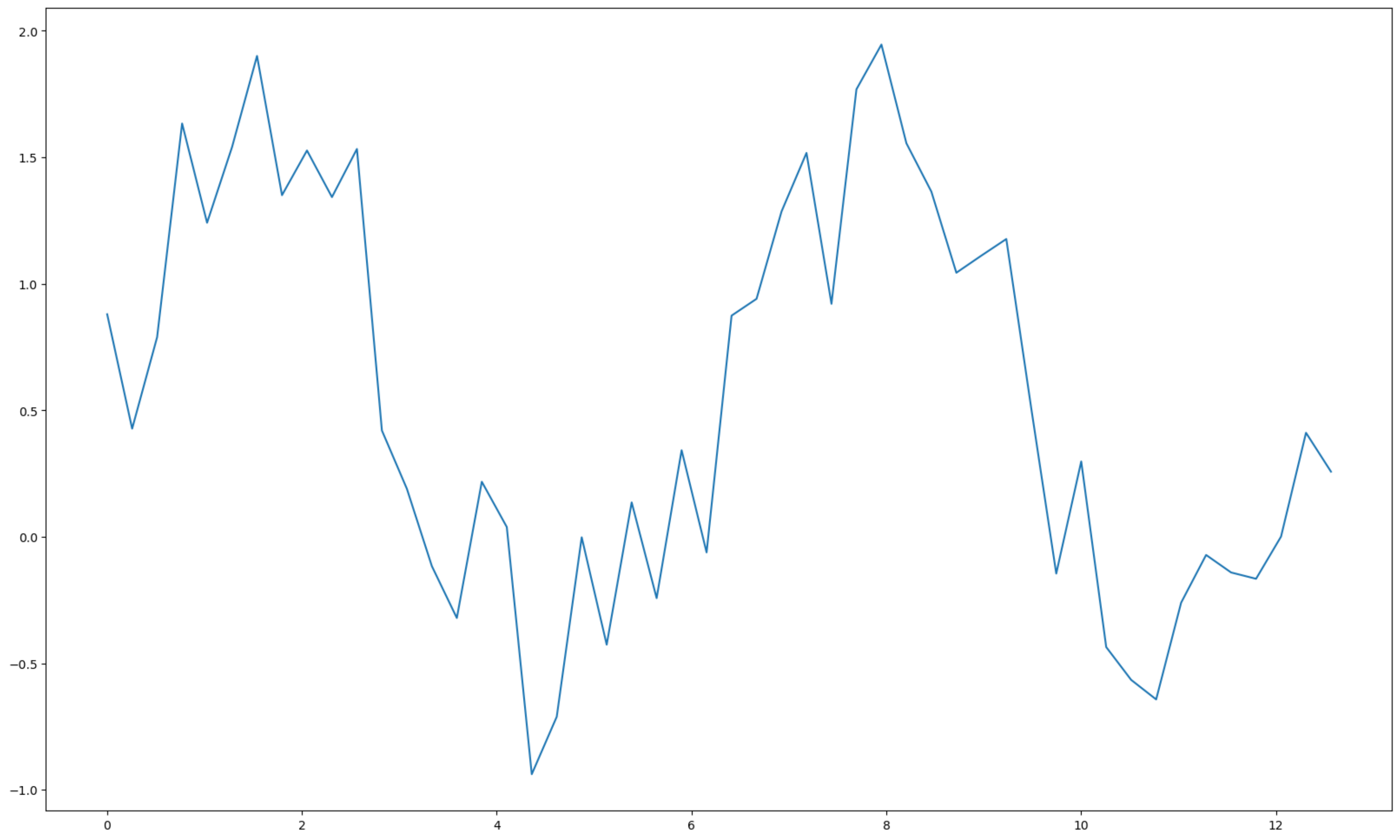
From the chart, it’s clear that the “elbow” is somewhere around the k value of 4. So, let’s run k-means with 4 clusters.



K-MEANS CLUSTERS FOR SPRING_WEAT IN RM 111, MAPLE CREEK



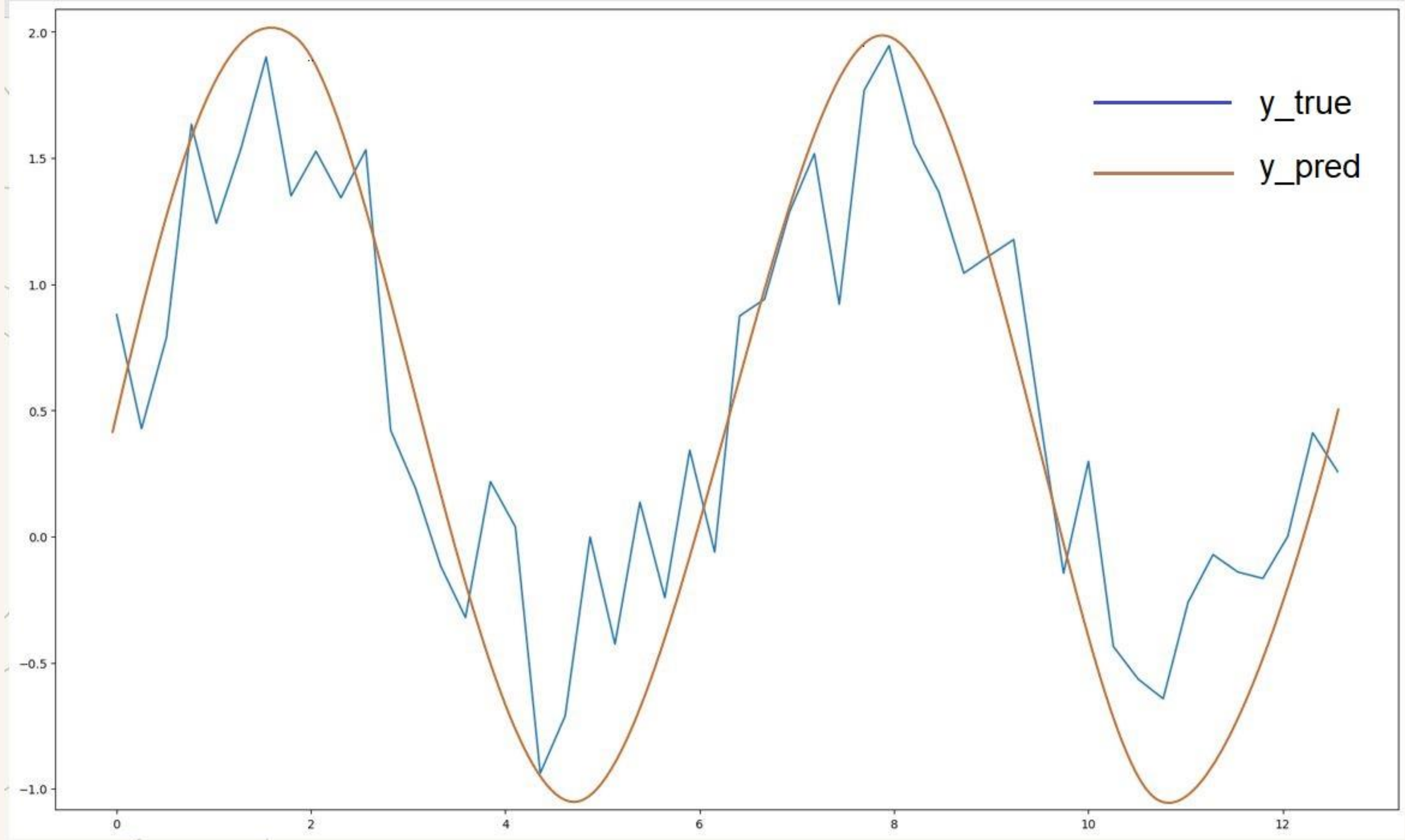
TO FIND MEA GENERATE DATA ADDED WITH NOISE



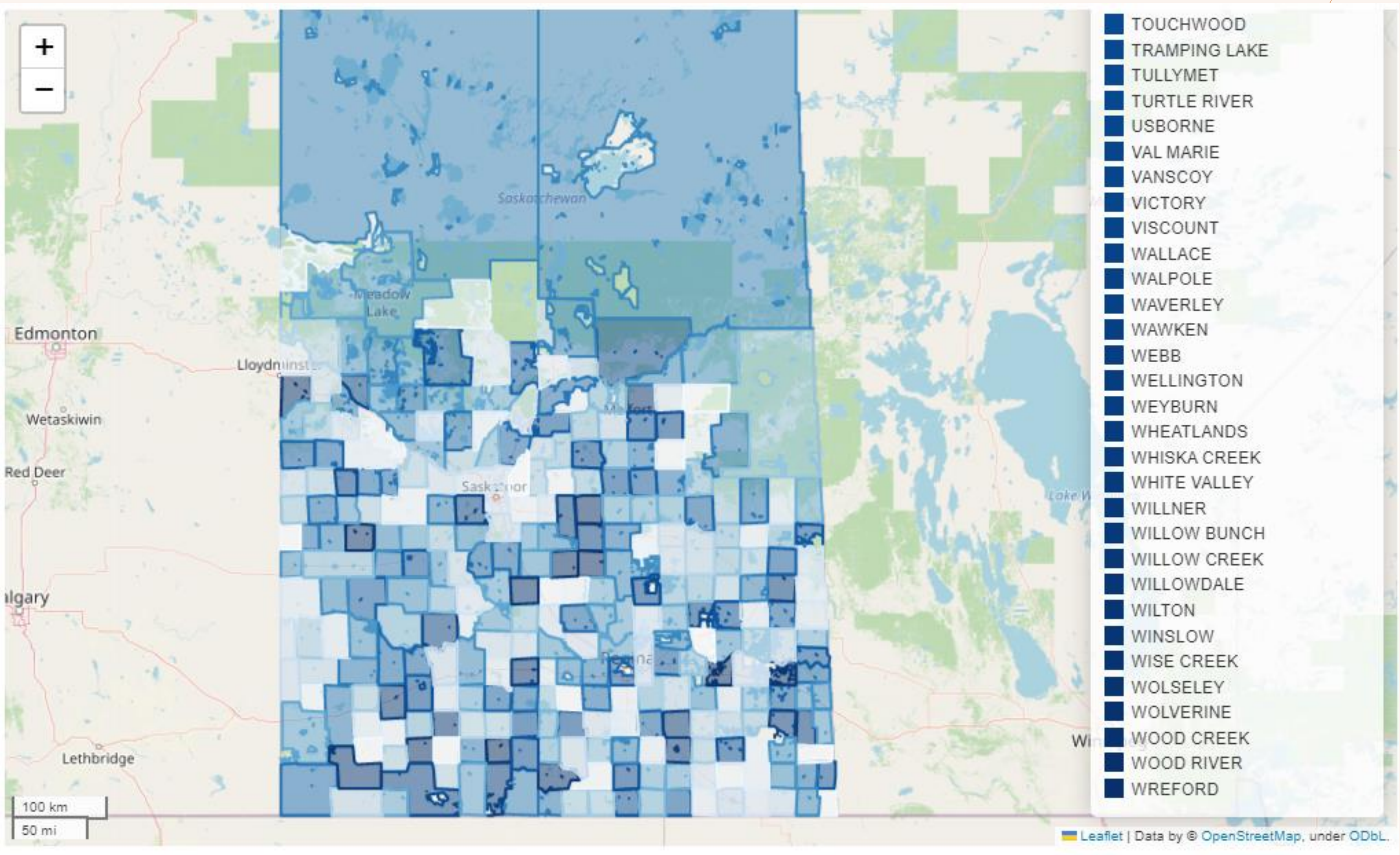
ASSUME THAT WE HAVE MADE A MODEL TO PREDICT THE Y VALUES FOR EACH X IN OUR DATASET.
SINUSOIDAL DATA WITH NOISE + PREDICTION

The model output with data

The mean absolute error is: 0.48

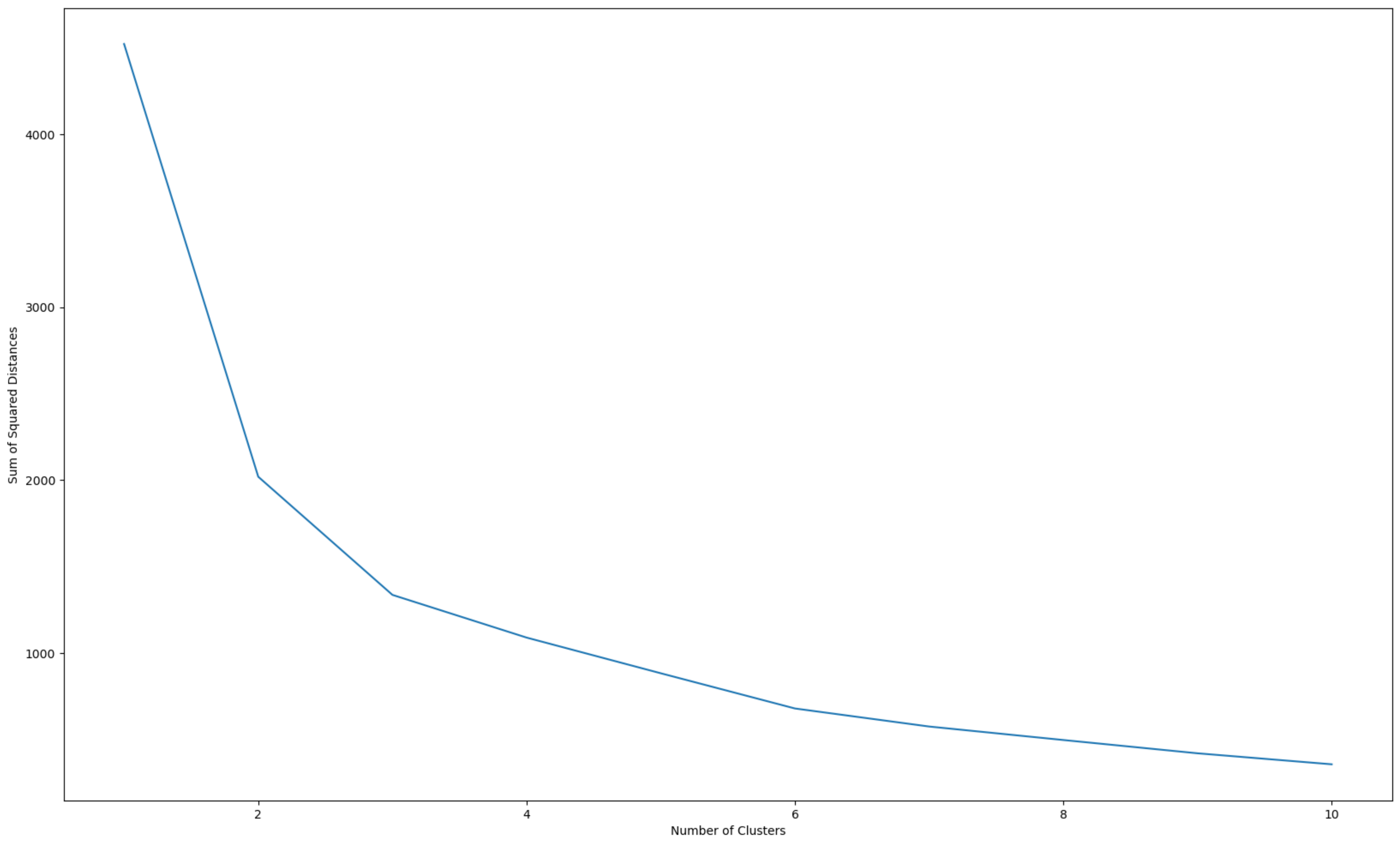


DATA MODELING FOR SK 622 RMS

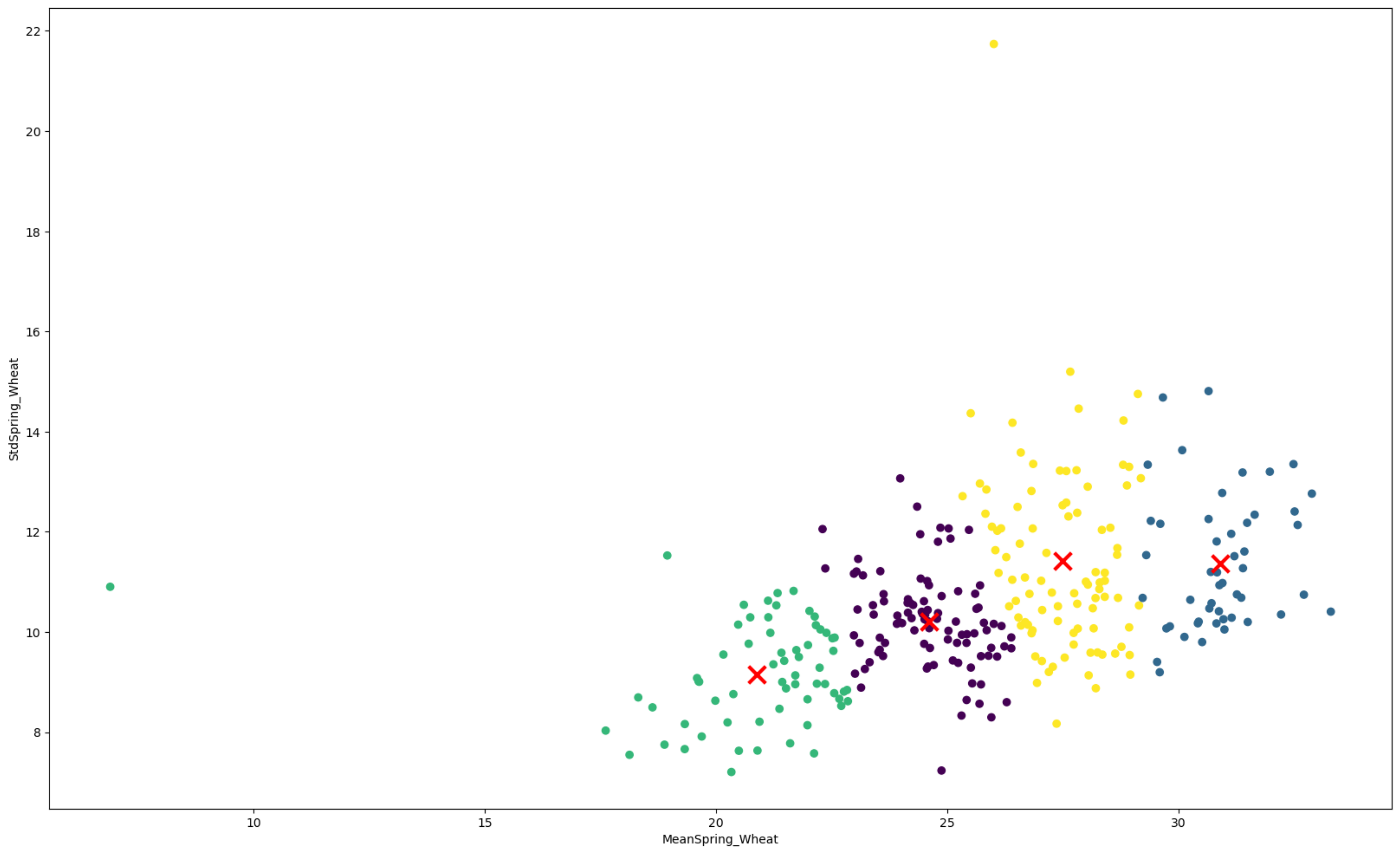

+
−
TOUCHWOOD
TRAMPING LAKE
TULLYMET
TURTLE RIVER
USBORNE
VAL MARIE
VANSCOY
VICTORY
VISCOUNT
WALLACE
WALPOLE
WAVERLEY
WAWKEN
WEBB
WELLINGTON
WEYBURN
WHEATLANDS
WHISKA CREEK
WHITE VALLEY
WILLNER
WILLOW BUNCH
WILLOW CREEK
WILLOWDALE
WILTON
WINSLOW
WISE CREEK
WOLSELEY
WOLVERINE
WOOD CREEK
WOOD RIVER
WREFORD
Saskatchewan
Edmonton
Wetaskiwin
Red Deer
Lethbridge
100 km
50 mi
Leaflet | Data by © OpenStreetMap, under ODbL.

K-MEANS CLUSTERING WITH SCIKIT-LEARN

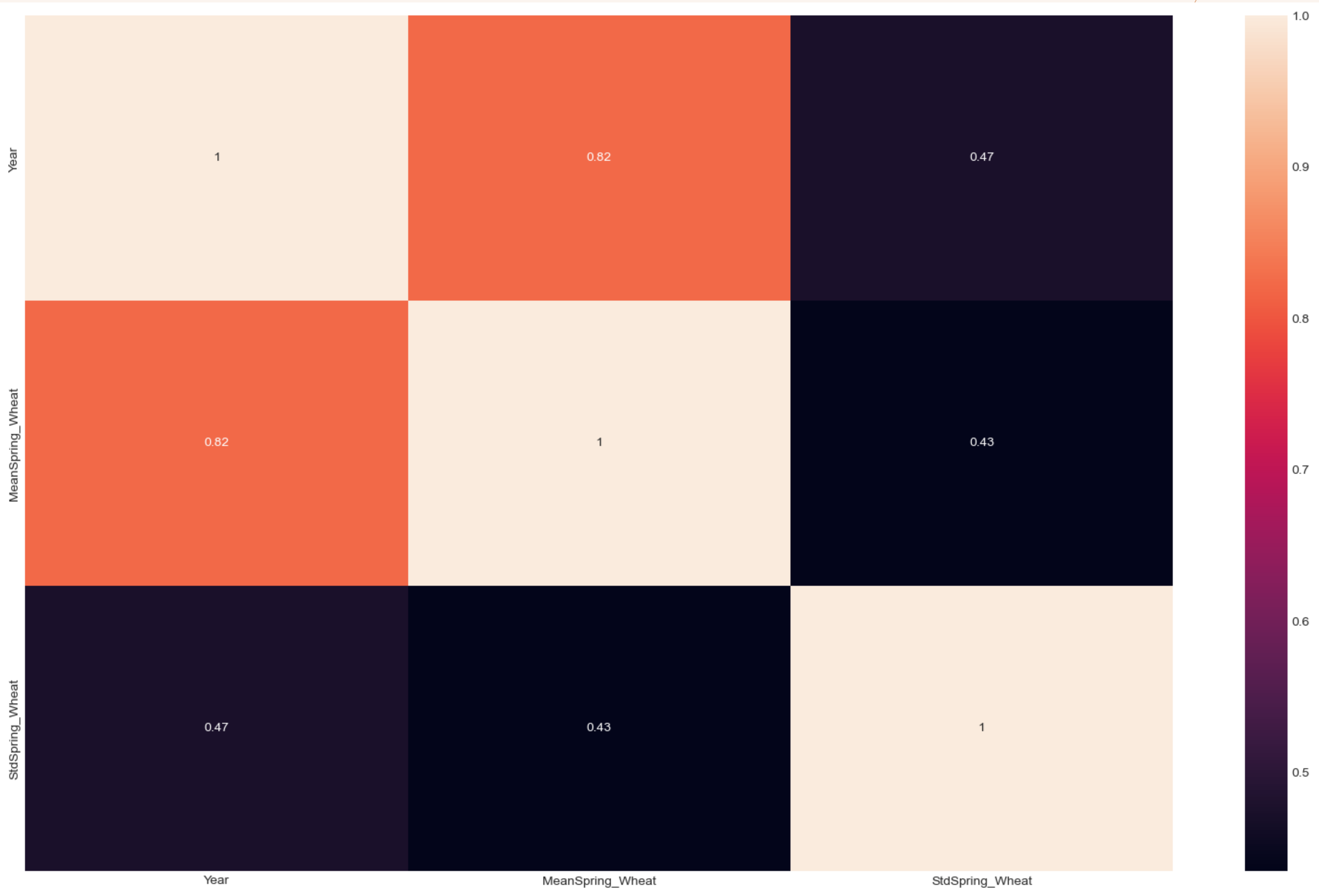
Found the “ Elbow”



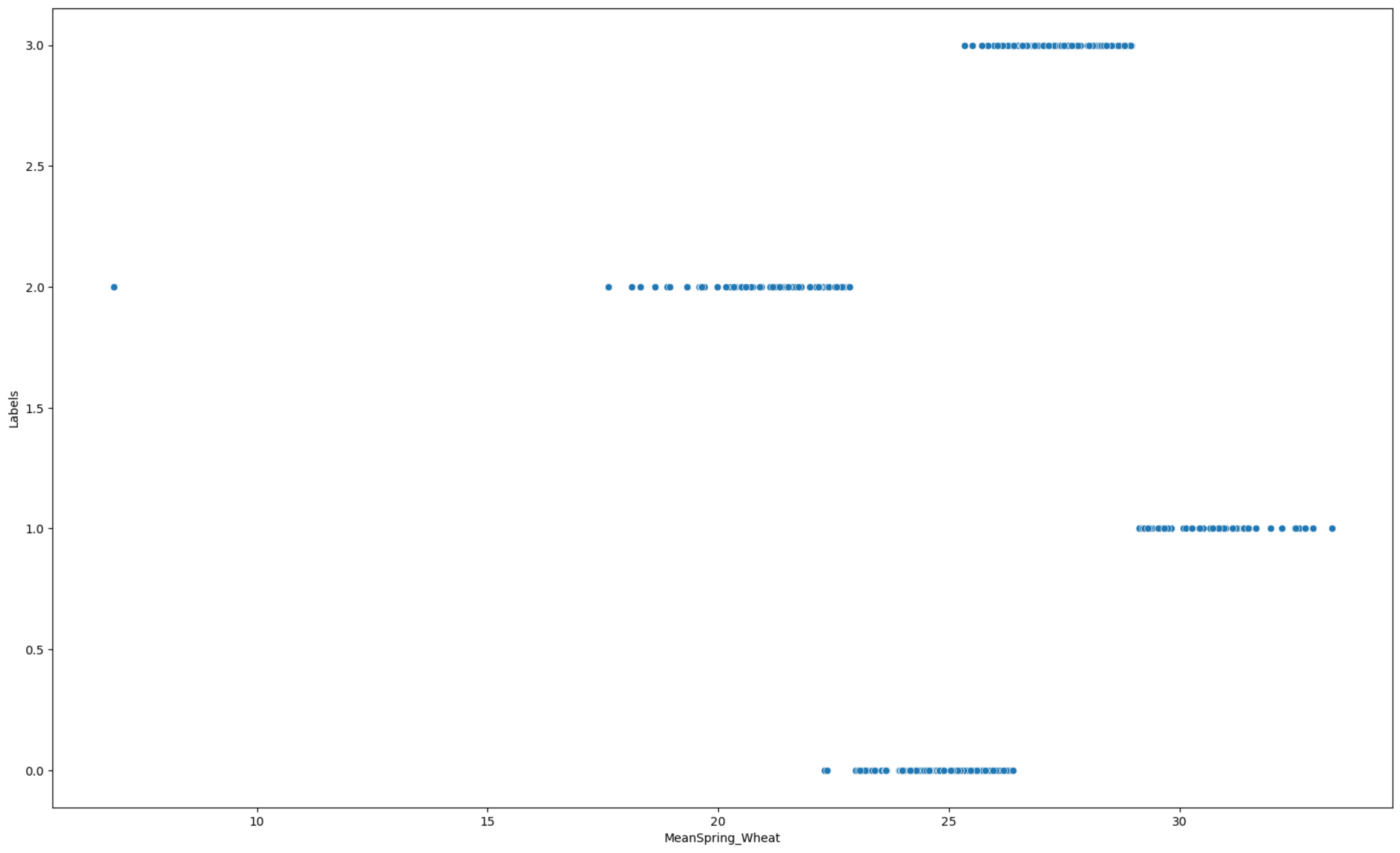
K-MEANS CLUSTERS IN SPRING_WEAT FORSK 622 RMS



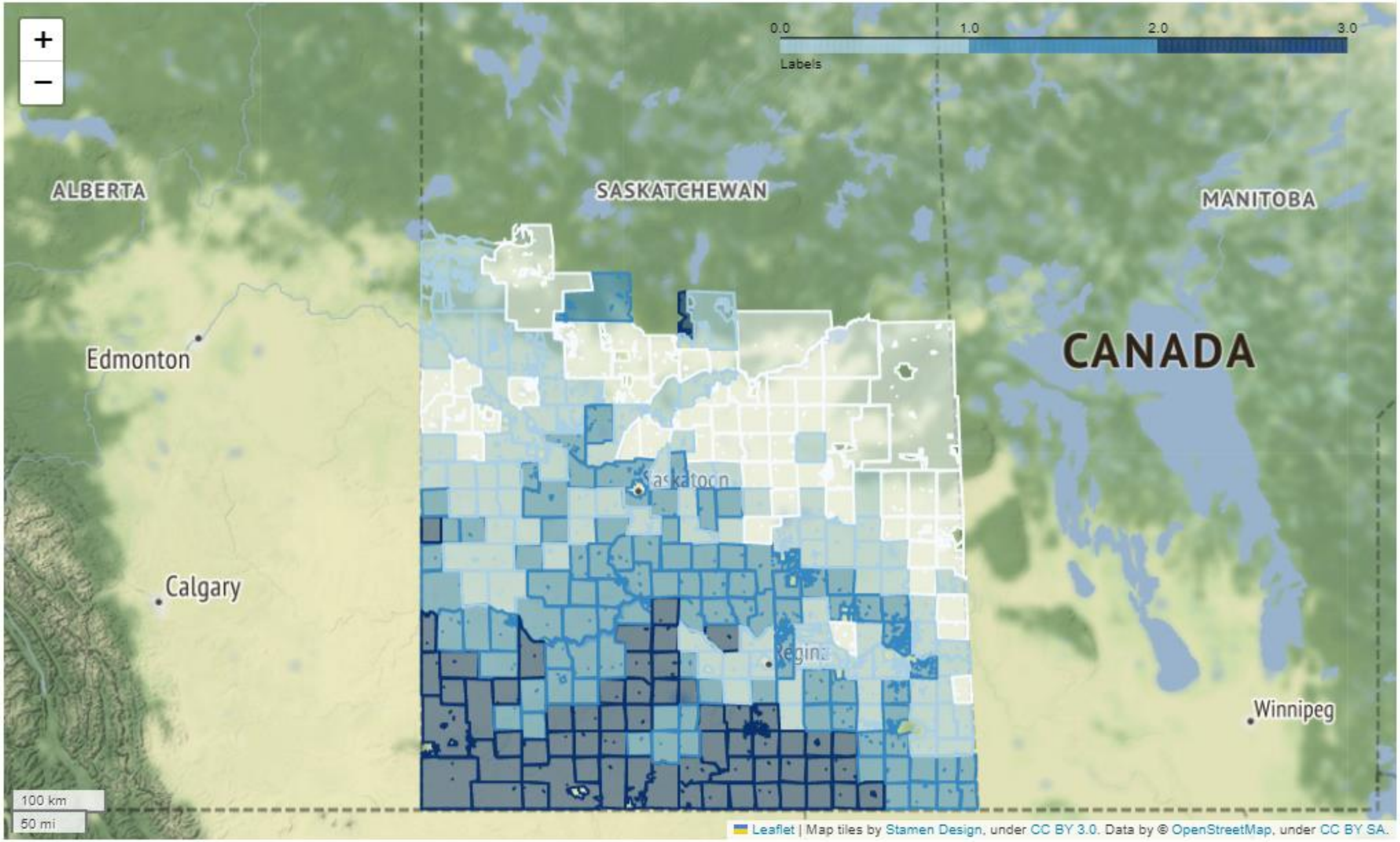
HEATMAP FOR SPRING_WEAT IN FOR SK 622 RMS



SCATTER PLOT FOR SK 622 RMS(EACH CLUSTERS)



Define the features to be used for clustering



RESULTS

The sizes of the four clusters are as follows: 0: n = 27.34, 1: n = 20.87; 2: n = 30.66, 3: n = 24.54. The majority of the RMs in the south are in Cluster 3. According to the unsupervised ML, the highest value is in cluster 2.In the north SK, there is an exceptional RM, which also indicates that it belongs to class 3.However, spring wheat's high production is demonstrated in South Saskechawan RMs.

THANK YOU