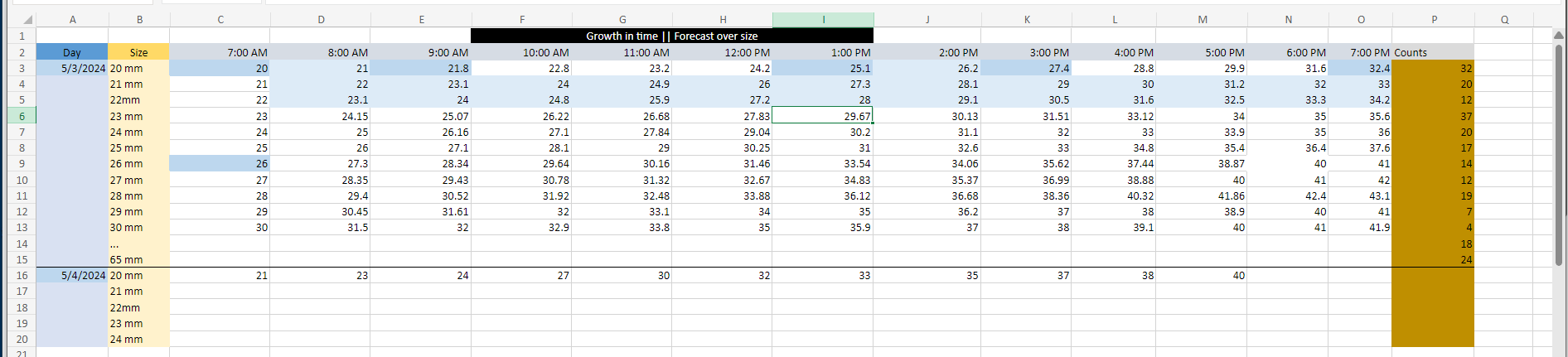
Growth rate of mushrooms

Mushrooms are living organisms but the interesting characteristic it has, is mushrooms are similar to plants when it comes to food intake and growths and become more like microorganisms when it comes to metabolism and environment.

Growth is the increment in size or other parameter while growth rate is growth in a specific time period. Mushroom growth and its growth rate can be calculated in different ways. Two different methods were used to estimate mushroom growth rate

1. Growth rate over full cycle analysis
2. Growth rate over pair differential technique

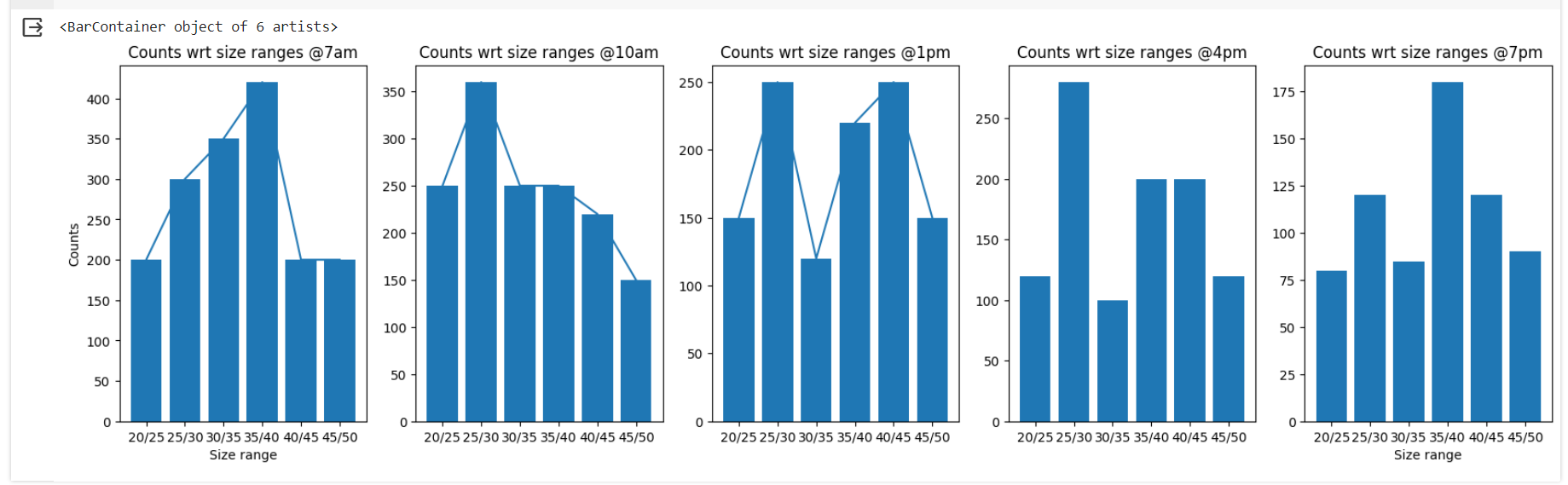
In both cases, growth was calculated in a time interval subtracting final size to the initial size within that time window. To estimate growth rate a time series data was used-



This is a data where individual mushroom sizes(20mm,21mm and so on) were monitored every hour and recorded what led to a time series growth data set.

Now the question is how to use this data to forecast mushroom growth ahead of time. Previously a different data set was made with counts of mushroom sizes over time and fed that to a model to predict count in 6 hours ahead given count changes every hour corresponding to each mushroom sizes as pickers carry on picking mushrooms.

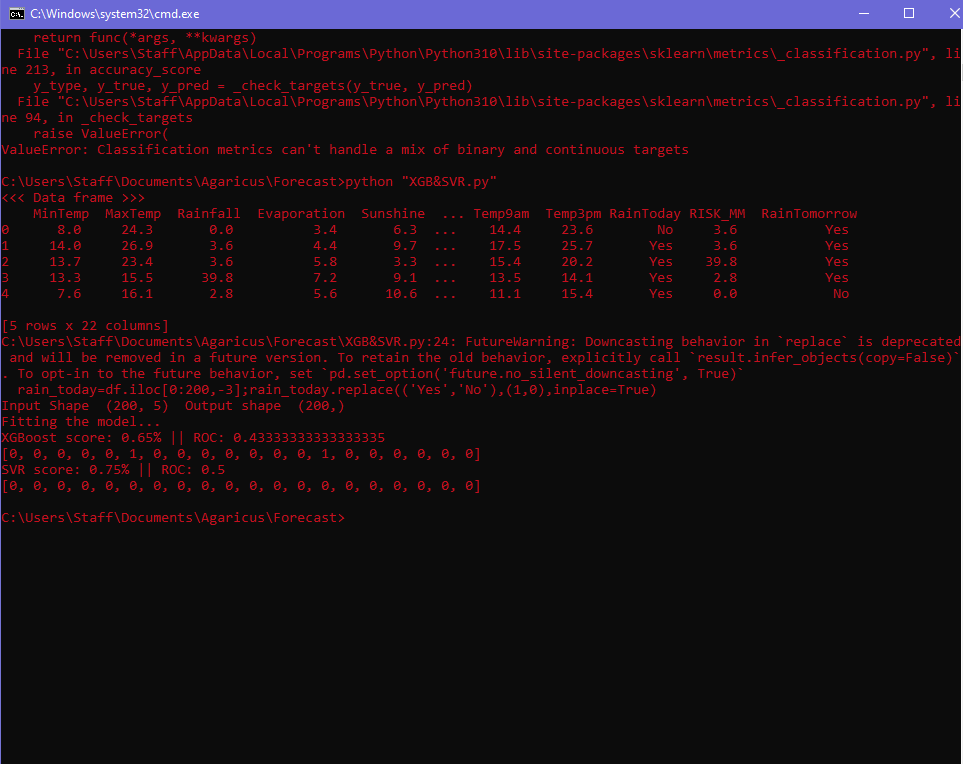
Below figure demonstrates how mushroom counts changes over time



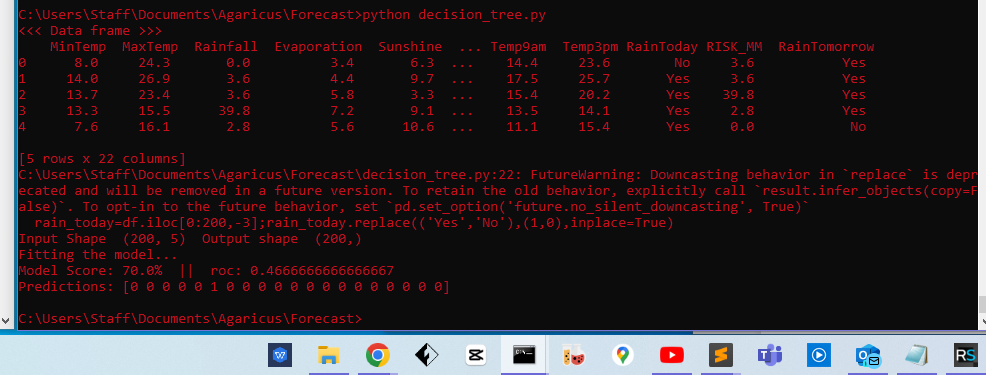
Some models previously tried:

To find a suitable model first similar dataset was searched. A weather forecast dataset matched the requirement. Several models were trained and evaluated over that dataset to see how better the models were performing. It seemed conventional models like Random Forest, Decision tree, Naive bayes, even boosting algorithms could not go beyond 75% accuracy.

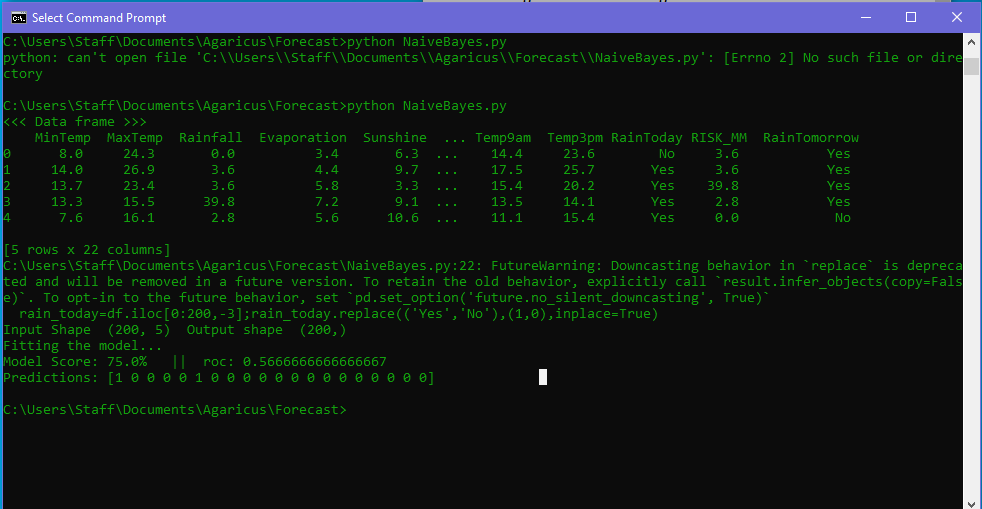
Selection of Models



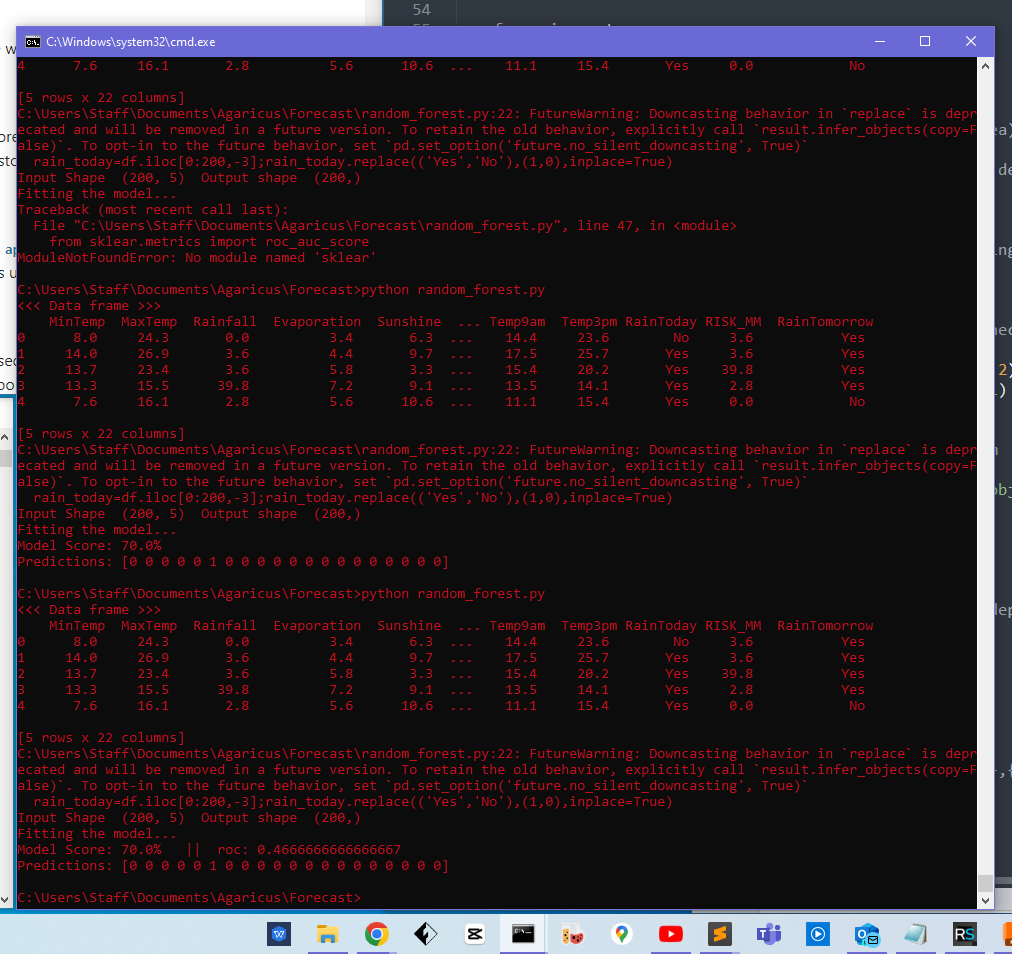
XGBoost



Decision Tree



Naive Bayes

Random forest

Now once satisfactory results were not found, Gaussian process regression was tried what gave 93% accuracy.

Inference:

It seemed regression might be a good choice to tackle this problem.

Restrictions:

The dataset used was medium in size. Bigger dataset could have given more inside of the suitability of the regression model. Actual mushroom dataset may differ in properties and inner statistics wrt this dataset

Step taken:

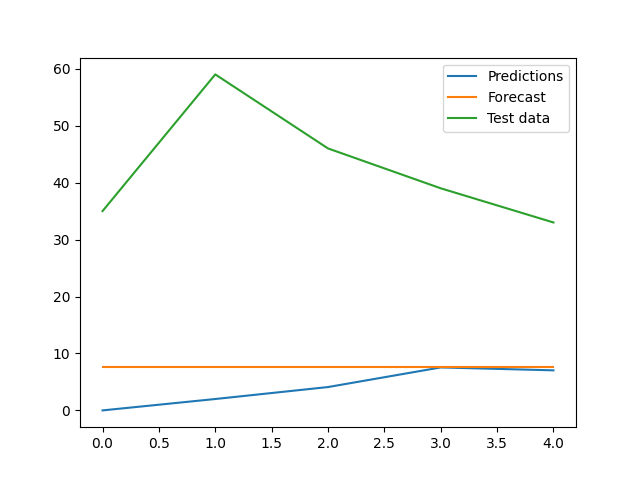
A fake dataset was made and stronger regression was searched.

Change of dataset and experiment of ARIMA:

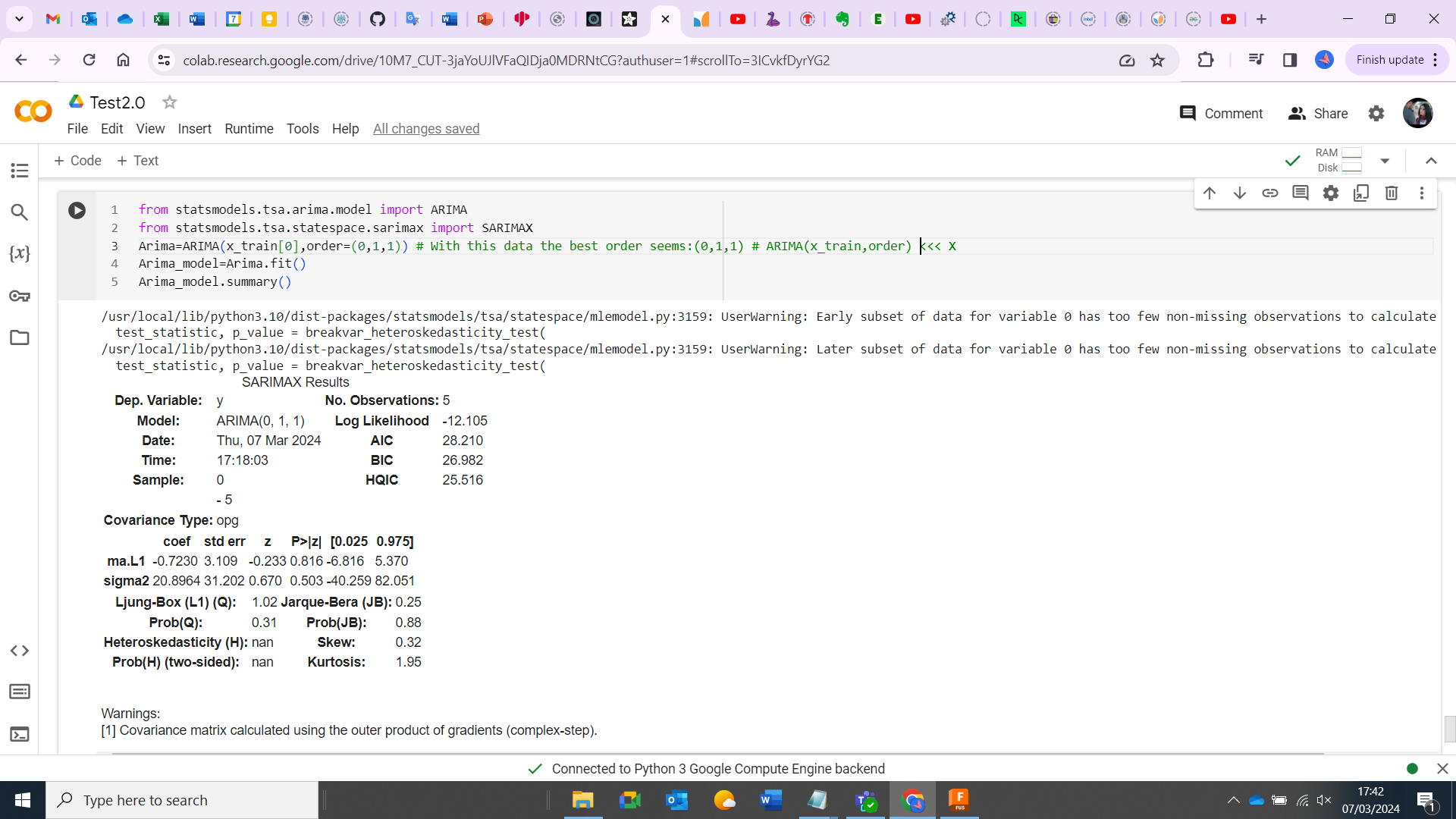
A test was performed whether a different dataset can do better forecast. A time series dataset with hourly growth was created and it acted as a fake dataset for the test performed afterwards.

In the beginning ARIMA (Autoregressive Integrated Moving Average) was tested and it forecast on 6 and 12 hours ahead of time.

Experimental results:



ARIMA forecast vs Test data comparison



Statistical Insight of the data

Inference:

1. ARIMA needs to be applied to individual sizes and it will forecast based on the input.

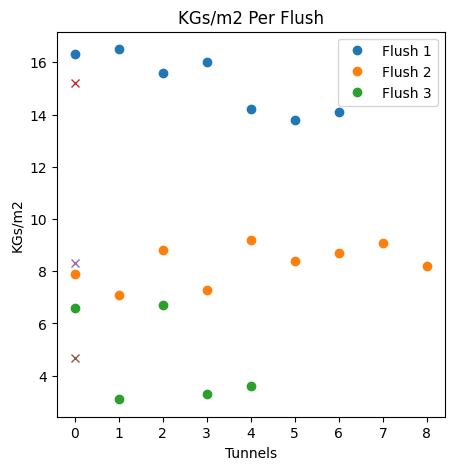
2. It cannot take multiple features one at a time e.g. 20mm time-series and 21mm time-series cannot be fed altogether but one by one after it does forecast one after another.

3. the feature input needs to be large, here hourly growth data was fed and that stood out not to be sufficient

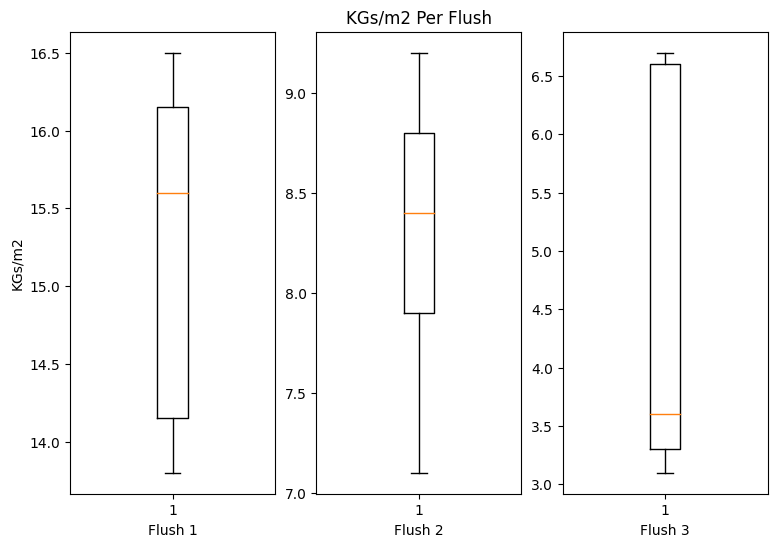
4. Data under test needs to be stationary and the data came out to be non-stationary so first the dataset was made differenced to make it stationary to feed into ARIMA

Steps taken:

1. Understanding growth rate and how it changes over time were taken under consideration and tests were performed
2. Before jumping to growth an overview was fetched though mushroom population over each tunnel. It seems population KGs/m2 does not vary that mush tunnel to tunnel.



KGs/m2 per tunnel variation



KGs/m2 variation per flush

Growth rate

It stands out that growth rate can be evaluated in two ways-

1. Growth rate over full cycle analysis (FCA)
2. Growth rate over pair differential technique (PD)

GR over FCA:

The growth was calculated over a whole 12 hours of time series data without differentiating the same into intervals.

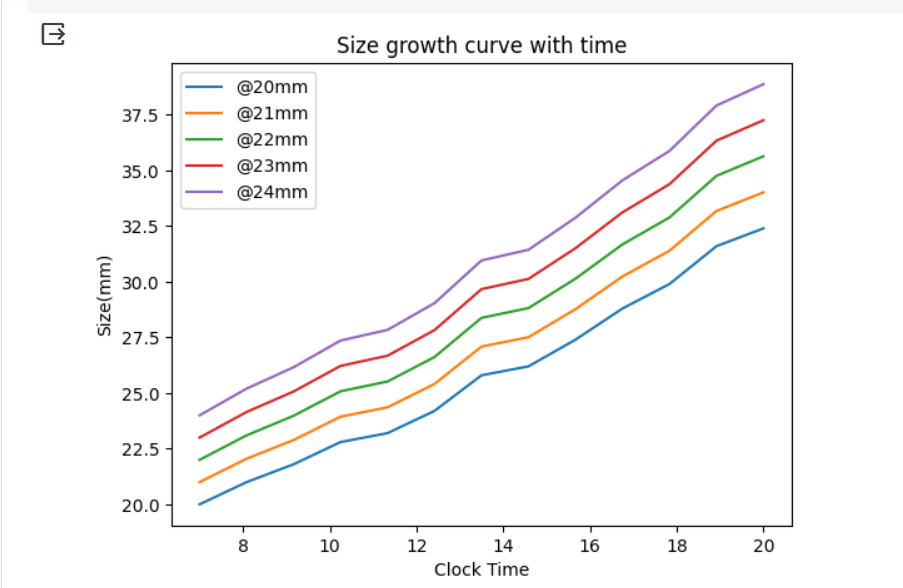
Growth=(Final size - Initial size); where Initial size= Size at 7 am; Final size= Size at 7 pm

Growth rate= Growth / Total time; where Total time =12 (hrs)

Results:

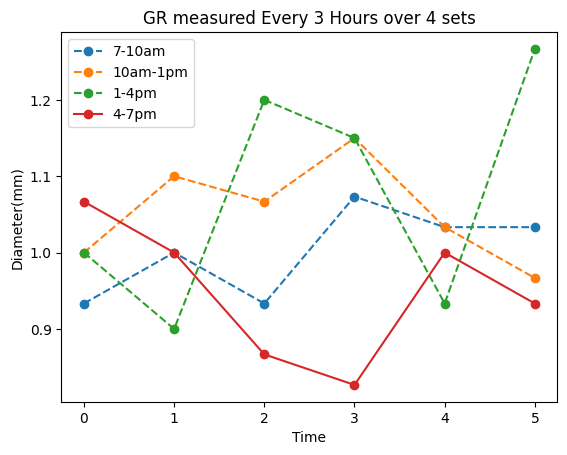
As expected, growth rate was incremental over time.

Note: The data was fake so not all the sizes will show exact same fluctuations and variations.



[GR\_PDTechnique.py](https://universityoflincoln-my.sharepoint.com/:u:/g/personal/akundu_lincoln_ac_uk/Eba5BgAW94RNvBLExkIFwtMBP_I5uG9lqe3ZlZiVHmyfmw?e=4DFenL) will take data from excel sheet and will give growth rate in different intervals dividing the data in periods.

\*\*\* Data Captured Every 1 Hour\*\*\*   
Statistical description of data over each available size   
DescribeResult(nobs=13, minmax=(20.0, 32.0), mean=25.78076923076923, variance=14.906474358974359, skewness=0.08036650092794528, kurtosis=-1.206167268164464) std: 3.8608903583207796  
DescribeResult(nobs=13, minmax=(21.0, 33.0), mean=27.04615384615385, variance=15.294358974358971, skewness=-0.01674513853568827, kurtosis=-1.2307721920962371) std: 3.910800298450302  
DescribeResult(nobs=13, minmax=(22.0, 34.2), mean=28.16923076923077, variance=16.510641025641025, skewness=0.0014235614958357675, kurtosis=-1.3098241717004027) std: 4.063328810918583  
DescribeResult(nobs=13, minmax=(23.0, 35.6), mean=29.383076923076924, variance=18.09163974358974, skewness=0.04441286492415426, kurtosis=-1.3244676456151794) std: 4.2534268235846895  
DescribeResult(nobs=13, minmax=(24.0, 36.0), mean=30.026153846153843, variance=15.016158974358971, skewness=-0.015584326615419785, kurtosis=-1.2051578270043217) std: 3.8750688993047557  
DescribeResult(nobs=13, minmax=(25.0, 37.6), mean=31.25, variance=16.787499999999994, skewness=0.016004190885846186, kurtosis=-1.2366152301990754) std: 4.097255178775176  
  
Growth over 12 hours for each sizes from 20 to 25mm-  
-----For each sizes from 20 to 25mm-----  
Growth: [12, 12, 12.2, 12.6, 12, 12.6]  
Growth rate: [1.0, 1.0, 1.0167, 1.05, 1.0, 1.05]  
  
  
 \*\*\* Data Captured Every 3 Hours\*\*\*   
Statistical description of data over each available size from 20 to 25mm  
DescribeResult(nobs=5, minmax=(20.0, 32.0), mean=25.879999999999995, variance=22.512, skewness=0.056528342106975556, kurtosis=-1.297549433005547) std: 4.744681232706787  
DescribeResult(nobs=5, minmax=(21.0, 33.0), mean=27.060000000000002, variance=22.518, skewness=-0.04234164561032463, kurtosis=-1.3015165805005882) std: 4.7453134775270644  
DescribeResult(nobs=5, minmax=(22.0, 34.2), mean=28.120000000000005, variance=24.39200000000001, skewness=0.002512489357474378, kurtosis=-1.4045919221858005) std: 4.9388257713752175  
DescribeResult(nobs=5, minmax=(23.0, 35.6), mean=29.522, variance=25.837320000000005, skewness=-0.09004611147601829, kurtosis=-1.3800603074535875) std: 5.083042396045896  
DescribeResult(nobs=5, minmax=(24.0, 36.0), mean=30.060000000000002, variance=22.357999999999997, skewness=-0.03563939514522429, kurtosis=-1.283943043197049) std: 4.7284246848184015  
DescribeResult(nobs=13, minmax=(25.0, 37.6), mean=32.88846153846154, variance=14.14089743589743, skewness=-0.5712923294253125, kurtosis=-0.4374398797282022) std: 3.7604384632509853  
  
Growth over 12 hours for each sizes from 20 to 25mm-  
-----For each sizes from 20 to 25mm-----  
Growth: [12, 12, 12.2, 12.6, 12, 12.6]  
Growth rate: [1.0, 1.0, 1.0167, 1.05, 1.0, 1.05]  
  
  
 \*\*\* Data divided into Every 6 Hours\*\*\*   
Variance (difference) 4.13 4.932261904761903 Skewness (difference) 0.08274834000486235 0.0931242438739891  
Variance (difference) 4.916190476190477 4.38142857142857 Skewness (difference) 0.0849525865985446 0.047535068531629025  
Variance (difference) 4.716666666666666 5.0514285714285725 Skewness (difference) 0.061684561698199865 -0.2259828023169921  
Variance (difference) 5.093514285714286 5.460980952380949 Skewness (difference) 0.2133392418500716 -0.1336773645364087  
Variance (difference) 4.801047619047616 4.3757142857142854 Skewness (difference) 0.03525853772226335 0.07212987951207674  
Variance (difference) 4.80892857142857 5.359999999999999 Skewness (difference) -0.032499961663110165 -0.08773780040180913  
  
---Growth over 6 hours for each sizes from 20 to 25mm(in list)--- 7 am to 1 pm  
Growth: [5.8, 6.3, 6, 6.67, 6.2, 6]  
Growth rate: [0.9667, 1.05, 1.0, 1.1117, 1.0333, 1.0]  
---Growth over 6 hours for each sizes from 20 to 25mm(in list)--- 1 pm to 7 pm  
Growth: [6.2, 5.7, 6.2, 5.93, 5.8, 6.6]  
Growth rate: [1.0333, 0.95, 1.0333, 0.9883, 0.9667, 1.1]  
  
  
  
 \*\*\* Data divided into Every 3 Hours\*\*\*   
Variance (difference) 1.6200000000000037 1.6424999999999967 Skewness (difference) 0.23408475365417544 0.16847151077905306  
Variance (difference) 2.030000000000001 1.3533333333333322 Skewness (difference) 0.19322827333053283 0.11881662569208375  
Variance (difference) 1.995833333333333 2.4866666666666672 Skewness (difference) -0.13130472159859916 7.846292258686524e-15  
Variance (difference) 2.36286666666667 2.411358333333327 Skewness (difference) 0.5802101365845006 0.4432023652042202  
Variance (difference) 1.8563666666666658 1.4425000000000001 Skewness (difference) 0.19581911612621147 0.058568078183191326  
Variance (difference) 1.663958333333333 2.4366666666666625 Skewness (difference) -0.06906992918010263 0.10837856616671766  
  
---Growth over 3 hours for each sizes from 20 to 25mm(in list)--- 7 am to 10 am  
Growth: [2.8, 3, 2.8, 3.22, 3.1, 3.1]  
Growth rate: [0.9333, 1.0, 0.9333, 1.0733, 1.0333, 1.0333]  
---Growth over 3 hours for each sizes from 20 to 25mm(in list)--- 10 am to 1 pm  
Growth: [3.0, 3.3, 3.2, 3.45, 3.1, 2.9]  
Growth rate: [1.0, 1.1, 1.0667, 1.15, 1.0333, 0.9667]  
---Growth over 3 hours for each sizes from 20 to 25mm(in list)--- 1 pm to 4 pm  
Growth: [3.0, 2.7, 3.6, 3.45, 2.8, 3.8]  
Growth rate: [1.0, 0.9, 1.2, 1.15, 0.9333, 1.2667]  
---Growth over 3 hours for each sizes from 20 to 25mm(in list)--- 4 pm to 7 pm  
Growth: [3.2, 3, 2.6, 2.48, 3, 2.8]  
Growth rate: [1.0667, 1.0, 0.8667, 0.8267, 1.0, 0.9333]



GR over PD technique:

In this method, the dataset was divided into each hour and then growth rate was calculated using data from two pairs and comparting with others.

Growth= 8am size-7am size , 9am size-8am size and so on

Growth rate=mean(sum of Growth)

[GR\_PDTechnique.py](https://universityoflincoln-my.sharepoint.com/:u:/g/personal/akundu_lincoln_ac_uk/Eba5BgAW94RNvBLExkIFwtMBP_I5uG9lqe3ZlZiVHmyfmw?e=XQodIJ) will take data from excell and produce the following report

Results:

\*\*\* General growth rate considering 1 hour pair data\*\*\* ---Growth rate/ growth per hours over 12 hours for each sizes from 20 to 25mm(in list)--- 7 am to 7pm

Hour basis Growth rate:

20mm>>> [1, 0.8, 1.0, 0.9, 0.8, 1.3, 0.9, 0.9, 1.2, 0.95, 0.95, 1.3]

21mm>>> [1, 1.1, 0.9, 0.9, 1.1, 1.3, 0.8, 0.9, 1, 1.2, 0.8, 1]

22mm>>> [1.1, 0.9, 0.8, 1.1, 1.3, 0.8, 1.1, 1.4, 1.1, 0.9, 0.8, 0.9]

23mm>>> [1.15, 0.92, 1.15, 0.46, 1.15, 1.84, 0.46, 1.38, 1.61, 0.88, 1, 0.6]

24mm>>> [1, 1.16, 0.94, 0.74, 1.2, 1.16, 0.9, 0.9, 1, 0.9, 1.1, 1],

25mm>>>[1, 1.1, 1.0, 0.9, 1.25, 0.75, 1.6, 0.4, 1.8, 0.6, 1.0, 1.2]]

Differential GR for each size:

[1.0, 1.0, 1.02, 1.05, 1.0, 1.05]

