1. IMPORTING THE DATASET AND NECESSARY LIBRARIES

DATASET SOURCE: https://data.gov.in/catalog/variety-wise-daily-market-prices-data-wheat

Done by: Aditi R, 2 MDS [2248023]

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
data=pd.read_excel(r"C:\Users\Aditi Ravi\Downloads\Wheat_2023.xlsx")
#Importing the dataset as a dataframe and storing it as data
```

2. UNDERSTANDING THE DATA

In [3]:
```
data.head()
```

Out[3]:

	state	district	market	commodity	variety	arrival_date	min_price	max_price	modal
0	Bihar	Muzaffarpur	Muzaffarpur	Wheat	147 Average	2023-01-03	1950.0	2800.0	:
1	Bihar	Muzaffarpur	Muzaffarpur	Wheat	147 Average	2023-01-04	2000.0	2800.0	:
2	Bihar	Muzaffarpur	Muzaffarpur	Wheat	147 Average	2023-01-05	1950.0	2800.0	:
3	Bihar	Muzaffarpur	Muzaffarpur	Wheat	147 Average	2023-01-09	2000.0	2900.0	:
4	Bihar	Muzaffarpur	Muzaffarpur	Wheat	147 Average	2023-01-10	2000.0	2900.0	:

In [4]:
```
data.tail()
```

Out[4]:

	state	district	market	commodity	variety	arrival_date	min_price	max_price	modal_pri
7630	West Bengal	Uttar Dinajpur	Raiganj	Wheat	Local	2023-01-12	2700.0	2800.0	275(
7631	West Bengal	Uttar Dinajpur	Raiganj	Wheat	Local	2023-01-13	2700.0	2800.0	275(
7632	West Bengal	Uttar Dinajpur	Raiganj	Wheat	Local	2023-01-14	2700.0	2800.0	275(
7633	West Bengal	Uttar Dinajpur	Raiganj	Wheat	Local	2023-01-15	2700.0	2800.0	275(
7634	West Bengal	Uttar Dinajpur	Raiganj	Wheat	Local	2023-01-16	2700.0	2800.0	275(

In [5]:

```
data.describe()
```

Out[5]:

	min_price	max_price	modal_price
count	7619.000000	7612.000000	7634.000000
mean	2536.964694	2763.356280	2652.908174
std	445.122096	679.814203	479.184000
min	245.000000	680.000000	660.000000
25%	2450.000000	2605.000000	2550.000000
50%	2550.000000	2725.000000	2650.000000
75%	2660.000000	2850.000000	2750.000000
max	34000.000000	38000.000000	38000.000000

In [6]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7635 entries, 0 to 7634
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   state         7635 non-null   object
 1   district      7635 non-null   object
 2   market        7635 non-null   object
 3   commodity     7635 non-null   object
 4   variety       7635 non-null   object
 5   arrival_date  7635 non-null   datetime64[ns]
 6   min_price     7619 non-null   float64
 7   max_price     7612 non-null   float64
 8   modal_price   7634 non-null   float64
 9   update_date   7635 non-null   datetime64[ns]
dtypes: datetime64[ns](2), float64(3), object(5)
memory usage: 596.6+ KB
```

In [7]:
```
#total number of null values in each column
data.isnull().sum()
```

Out[7]:
```
state           0
district        0
market          0
commodity       0
variety         0
arrival_date    0
min_price      16
max_price      23
modal_price     1
update_date     0
dtype: int64
```

3. CLEANING THE DATA

In [8]:
```
#filling in null values
data = data.fillna(data.mode().iloc[0])
```

In [9]:
```
#checking to see if the total number of null values in each column is zero
data.isnull().sum()
```

Out[9]:
```
state           0
district        0
market          0
commodity       0
variety         0
arrival_date    0
min_price       0
max_price       0
modal_price     0
update_date     0
dtype: int64
```

In [10]:
```
#checking to see if there are duplicated values
data.duplicated().sum()
```

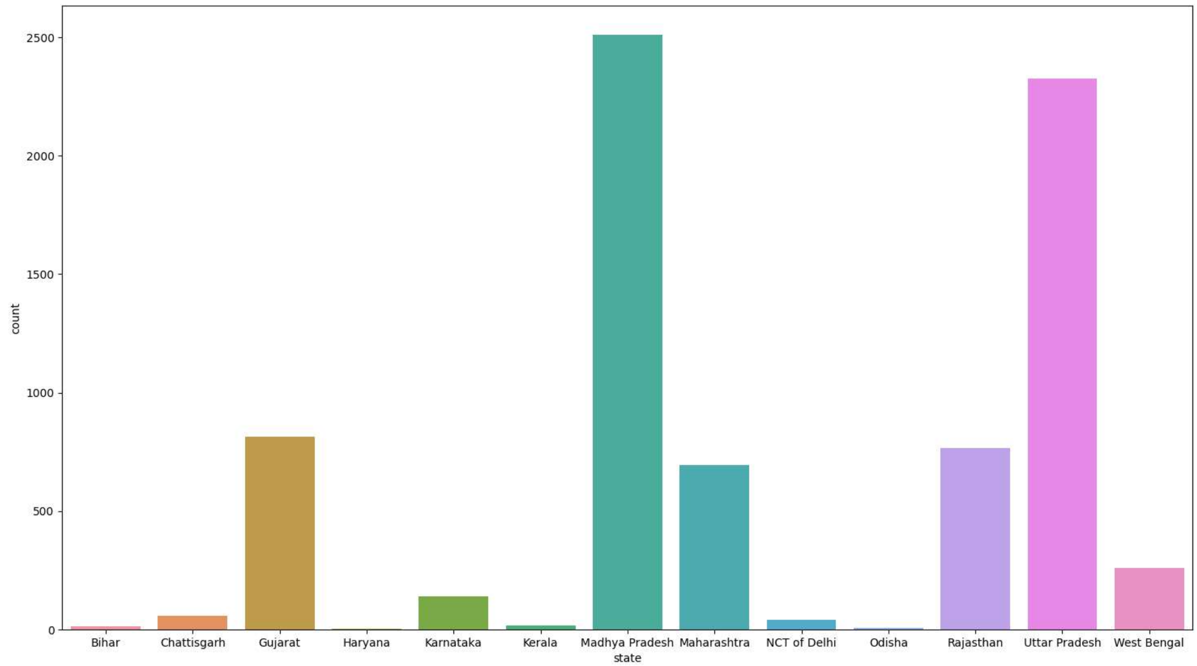
Out[10]: 0

4. EXPLORATORY DATA ANALYSIS

In [11]:
```
#Generating a count plot for the states
print(data['state'].value_counts())
plt.figure(figsize=(18,10))
sns.countplot(x='state', data=data)
```

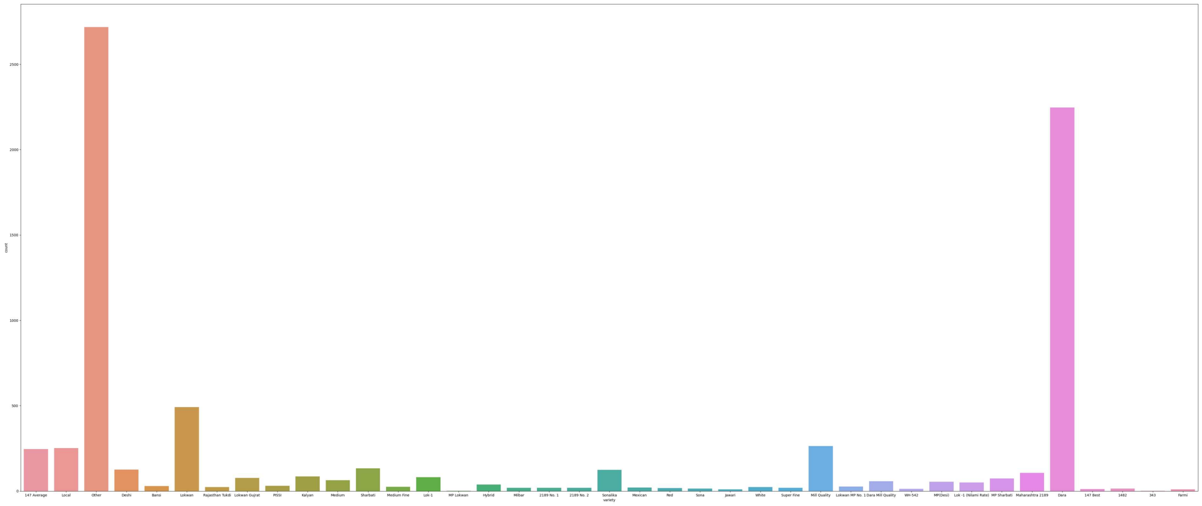
```
Madhya Pradesh    2510
Uttar Pradesh     2325
Gujarat            812
Rajasthan          766
Maharashtra        694
West Bengal        260
Karnataka          138
Chattisgarh         56
NCT of Delhi        39
Kerala              15
Bihar               14
Odisha               5
Haryana              1
Name: state, dtype: int64
```

Out[11]: `<AxesSubplot: xlabel='state', ylabel='count'>`



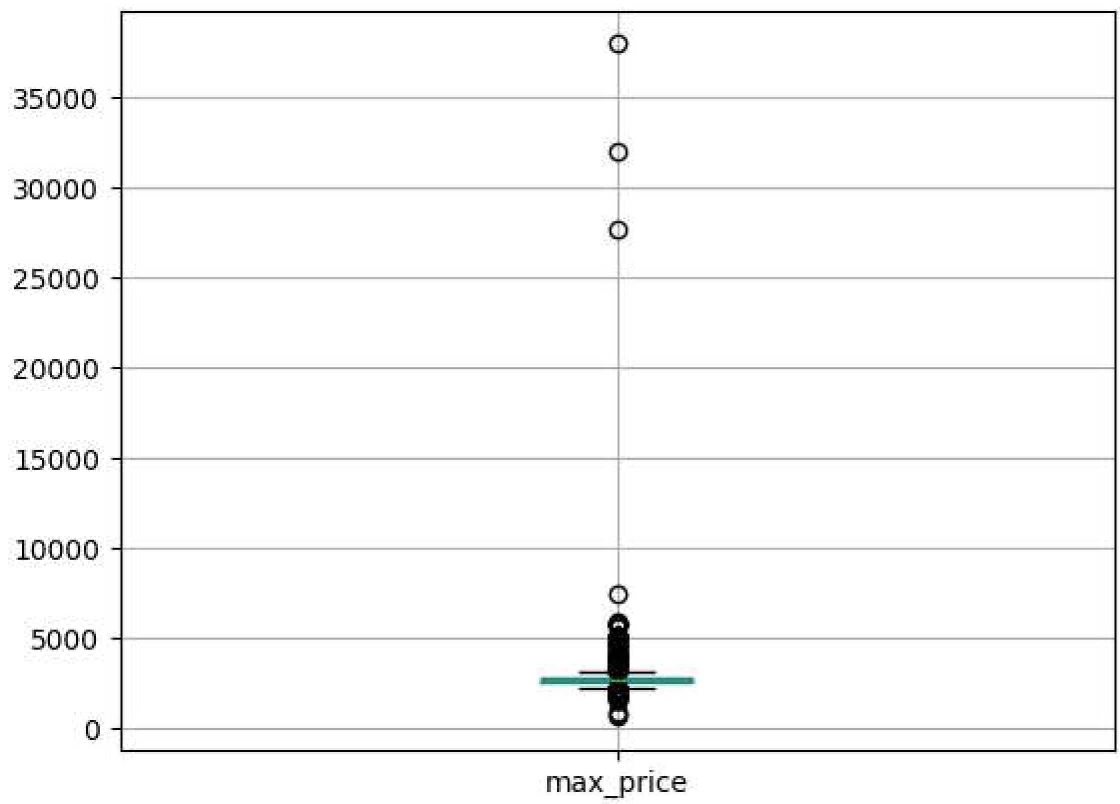
In [12]:
```python
#Generating a count plot for the districts
plt.figure(figsize=(60,25))
sns.countplot(x='variety', data=data)
```

Out[12]: `<AxesSubplot: xlabel='variety', ylabel='count'>`



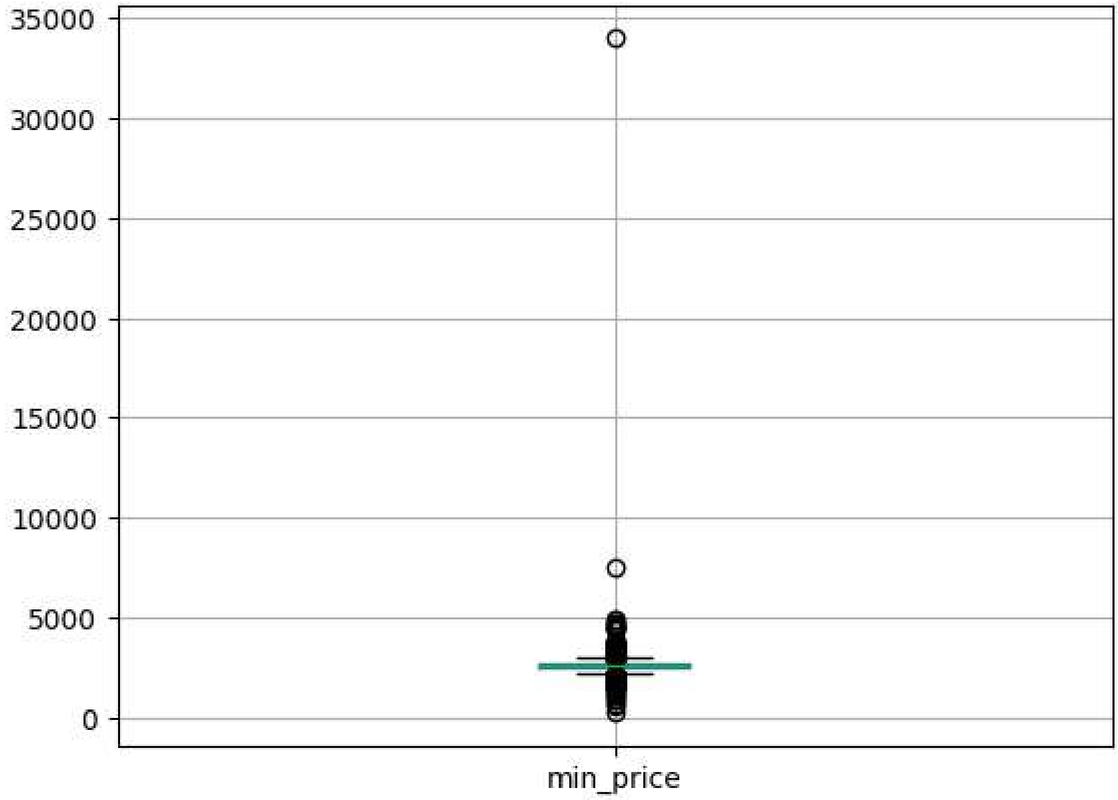
In [54]:
```
#Boxplot for maximum price
data[['max_price']].boxplot()
```

Out[54]: <AxesSubplot: >



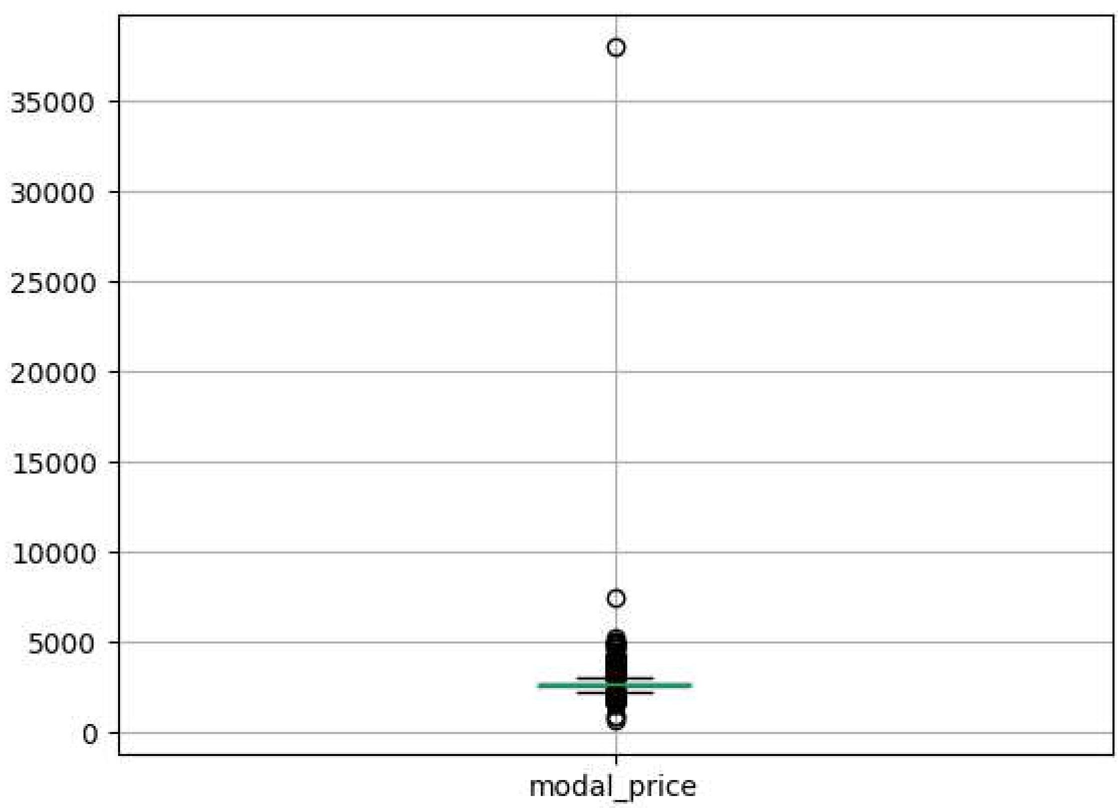
In [13]:
```
#Boxplot for minimum price
data[['min_price']].boxplot()
```

Out[13]: <AxesSubplot: >

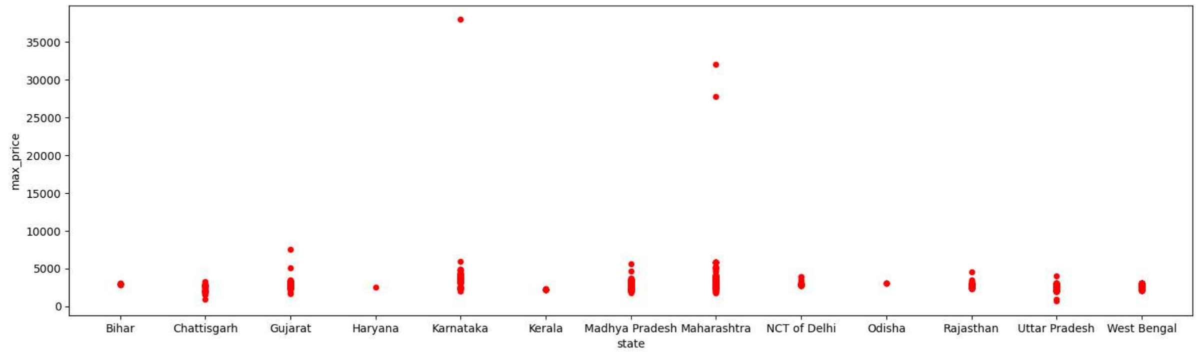


In [56]:
```
#Boxplot for modal price
data[['modal_price']].boxplot()
```

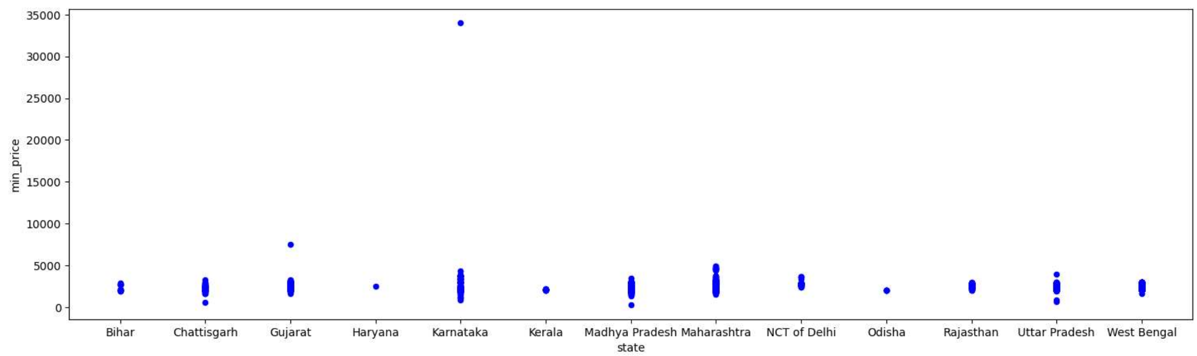
Out[56]: <AxesSubplot: >



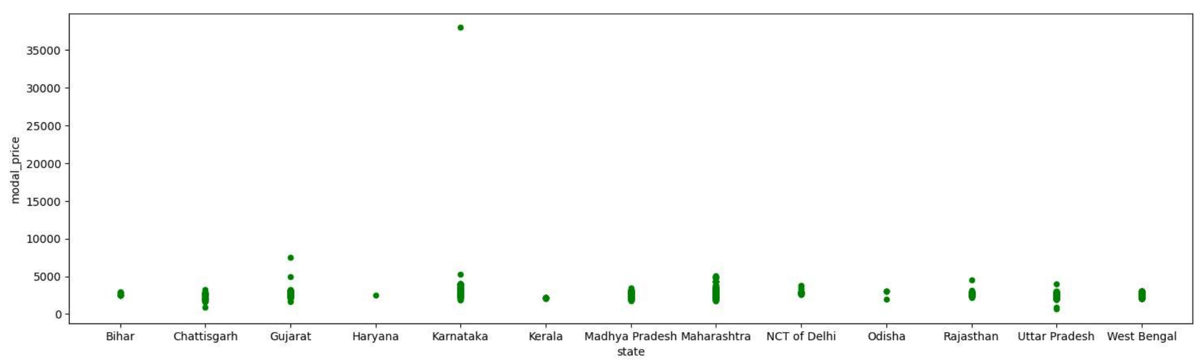
In [14]:
```python
#The scatter plot for maximum price
data.plot.scatter(x = 'state', y = 'max_price', s = 20, c='red', figsize=(18,5));
```



In [15]:
```python
#The scatter plot for minimum price
data.plot.scatter(x = 'state', y = 'min_price', s = 20, c='blue', figsize=(18,5));
```



In [16]:
```python
#The scatter plot for modal price
data.plot.scatter(x = 'state', y = 'modal_price', s = 20, c='green', figsize=(18,5)
```



In [17]:
```python
#To find the state with highest price
costly_state = data[data['max_price'] == data['max_price'].max()]
print("Inference: The state with highest wheat price is: ",costly_state.state.value
```

```
Inference: The state with highest wheat price is:  Karnataka and it is Rs. 38000.0
```

In [18]:
```python
#To find the Market with lowest price
least_price = data[data['min_price'] == data['min_price'].min()]
print("Inference: The state with lowest wheat price is: ",least_price.market.values
```

```
Inference: The state with lowest wheat price is:  Dewas and it is Rs. 245.0
```

In [19]:
```
#To find the variety with highest price
var = data[data['max_price'] == data['max_price'].max()]
print("Inference: The variety of wheat which is the the most priced is ",var.variet
```

```
Inference: The variety of wheat which is the the most priced is  Local and it is R
s. 38000.0
```

In [20]:
```
#To find the variety with lowest price
var = data[data['min_price'] == data['min_price'].min()]
print("Inference: The variety of wheat which is the the least priced is ",var.varie
```

```
Inference: The variety of wheat which is the the least priced is  Lokwan and it is
Rs. 245.0
```

CONCLUSION: We see that the Wheat sales in Madhya Pradesh is the highest and Karnataka sells it at the highest price. We conclude that our data is not normally distributed and it has outliers. We can further predict the evrage wheat price for the entire year.