Crop Production Analysis in India The Agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet. This paper presents a novel Business-to-Business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner. The objective of this project is to predict crop production and find important insights highlighting key indicators and metrics that influence crop production. # importing libraries import numpy as np import pandas as pd import matplotlib.pyplot as plt import seaborn as sns import warnings warnings.filterwarnings("ignore") # reading the data crop = pd.read_csv("Crop Production data.csv") pd.set_option("display.max_rows",1000) In [3]: crop.head(200) In [4] Out[4]: State_Name District_Name Crop_Year Season Crop Area Production **0** Andaman and Nicobar Islands **NICOBARS** 2000 1254.00 2000.00 Kharif Arecanut 1 Andaman and Nicobar Islands **NICOBARS** 2000 Kharif Other Kharif pulses 2.00 1.00 2 Andaman and Nicobar Islands **NICOBARS** 2000 Kharif Rice 102.00 321.00 2000 Whole Year 3 Andaman and Nicobar Islands **NICOBARS** 176.00 641.00 Banana 4 Andaman and Nicobar Islands **NICOBARS** 2000 Whole Year Cashewnut 720.00 165.00 2000 Whole Year 18168.00 65100000.00 5 Andaman and Nicobar Islands **NICOBARS** Coconut 6 Andaman and Nicobar Islands **NICOBARS** 2000 Whole Year Dry ginger 36.00 100.00 7 Andaman and Nicobar Islands **NICOBARS** 2000 Whole Year Sugarcane 1.00 2.00 8 Andaman and Nicobar Islands Sweet potato 15.00 **NICOBARS** 2000 Whole Year 5.00 Andaman and Nicobar Islands **NICOBARS** 2000 Whole Year Tapioca 40.00 169.00 10 Andaman and Nicobar Islands **NICOBARS** 2001 Kharif 1254.00 2061.00 Arecanut 11 Andaman and Nicobar Islands **NICOBARS** 2001 Kharif Other Kharif pulses 2.00 1.00 **12** Andaman and Nicobar Islands **NICOBARS** 2001 Kharif Rice 83.00 300.00 13 Andaman and Nicobar Islands **NICOBARS** 2001 Whole Year 719.00 192.00 Cashewnut 14 Andaman and Nicobar Islands 64430000.00 **NICOBARS** 2001 Whole Year 18190.00 Coconut 15 Andaman and Nicobar Islands **NICOBARS** 2001 Whole Year Dry ginger 46.00 100.00 **16** Andaman and Nicobar Islands **NICOBARS** 2001 Whole Year Sugarcane 1.00 1.00 Andaman and Nicobar Islands **NICOBARS** 2001 Whole Year Sweet potato 33.00 11.00 **18** Andaman and Nicobar Islands **NICOBARS** 2002 Kharif 189.20 510.84 Rice 19 Andaman and Nicobar Islands **NICOBARS** 2002 Whole Year 1258.00 2083.00 Arecanut Andaman and Nicobar Islands **NICOBARS** 2002 Whole Year 213.00 1278.00 Banana 21 Andaman and Nicobar Islands **NICOBARS** 2002 Whole Year Black pepper 63.00 13.50 22 Andaman and Nicobar Islands **NICOBARS** 2002 Whole Year 719.00 208.00 Cashewnut 2002 Whole Year 23 Andaman and Nicobar Islands **NICOBARS** Coconut 18240.00 67490000.00 Dry chillies 24 Andaman and Nicobar Islands 413.00 28.80 **NICOBARS** 2002 Whole Year 25 Andaman and Nicobar Islands **NICOBARS** 2002 Whole Year Dry ginger 47.30 133.00 **NICOBARS** 40.00 Andaman and Nicobar Islands 2002 Whole Year Sugarcane 5.00 27 Andaman and Nicobar Islands **NICOBARS** 2003 Kharif Rice 52.00 90.17 2003 Whole Year 1525.00 28 Andaman and Nicobar Islands **NICOBARS** Arecanut 1261.00 29 Andaman and Nicobar Islands **NICOBARS** 2003 Whole Year Banana 266.00 1763.00 **NICOBARS** 2003 Whole Year 30 Andaman and Nicobar Islands Black pepper 75.50 15.86 31 Andaman and Nicobar Islands **NICOBARS** 2003 Whole Year Cashewnut 717.00 208.50 32 Andaman and Nicobar Islands **NICOBARS** 2003 Whole Year Coconut 18284.74 68580000.00 Andaman and Nicobar Islands **NICOBARS** 2003 Whole Year Dry chillies 60.00 102.00 **NICOBARS 34** Andaman and Nicobar Islands 2003 Whole Year Dry ginger 102.00 326.40 35 Andaman and Nicobar Islands **NICOBARS** other oilseeds 2003 Whole Year 4.20 0.30 72.57 Andaman and Nicobar Islands **NICOBARS** 2004 Kharif Rice 52.94 37 Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year Arecanut 1264.70 805.85 **38** Andaman and Nicobar Islands **NICOBARS** 404.12 2004 Whole Year 66.42 Banana Black pepper 39 Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year 85.00 15.86 197.90 40 Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year 683.50 Cashewnut Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year Coconut 18394.70 52380000.00 42 Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year Dry chillies 60.00 145.20 43 Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year Dry ginger 118.00 366.00 0.11 44 Andaman and Nicobar Islands **NICOBARS** 2004 Whole Year 1.00 other oilseeds 45 Andaman and Nicobar Islands **NICOBARS** 2005 Kharif Rice 2.09 12.06 **NICOBARS** Andaman and Nicobar Islands 2005 Whole Year Arecanut 795.67 NaN 47 Andaman and Nicobar Islands **NICOBARS** 2005 Whole Year Banana 22.00 582.00 0.10 48 Andaman and Nicobar Islands **NICOBARS** 2005 Whole Year Black pepper 41.00 Andaman and Nicobar Islands **NICOBARS** 2005 Whole Year Cashewnut 450.00 55.80 **NICOBARS 50** Andaman and Nicobar Islands 2005 Whole Year Coconut 13876.57 52030000.00 51 Andaman and Nicobar Islands **NICOBARS** 2005 Whole Year Dry chillies 17.00 NaN **52** Andaman and Nicobar Islands **NICOBARS** 2005 Whole Year Turmeric 10.00 70.00 53 Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year Arecanut 896.00 477.60 **NICOBARS** 976.00 **54** Andaman and Nicobar Islands 2006 Whole Year 198.00 Banana 55 Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year Black pepper 22.16 0.68 **56** Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year Cashewnut 450.00 85.13 Andaman and Nicobar Islands **NICOBARS** Whole Year Coconut 14358.00 59370000.00 **NICOBARS** Dry chillies **58** Andaman and Nicobar Islands 2006 Whole Year 3.00 9.00 59 Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year Dry ginger 6.00 21.00 **NICOBARS** 0.50 Andaman and Nicobar Islands 2006 Whole Year Sugarcane 0.20 Sweet potato **61** Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year 18.00 86.00 **62** Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year 52.03 450.00 Tapioca 63 Andaman and Nicobar Islands **NICOBARS** 2006 Whole Year Turmeric 2.50 1.03 64 Andaman and Nicobar Islands **NICOBARS** 2010 Rice 3.50 10.00 Autumn Andaman and Nicobar Islands **NICOBARS** 2010 Autumn Sugarcane 13.40 41.75 66 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Arecanut 944.00 1610.00 67 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Black pepper 23.00 8.50 1000.50 260.50 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Cashewnut 69 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Dry chillies 12.00 25.00 **70** Andaman and Nicobar Islands **NICOBARS** 2010 9.64 Rabi Dry ginger 7.00 71 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Maize 3.84 18.22 72 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Moong(Green Gram) 1.50 1.10 73 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi 22.00 208.00 Sweet potato 74 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Turmeric 2.00 0.50 75 Andaman and Nicobar Islands **NICOBARS** 2010 Rabi Urad 1.50 1.16 **NICOBARS** 3425.00 **76** Andaman and Nicobar Islands 2010 Whole Year Banana 593.00 77 Andaman and Nicobar Islands **NICOBARS** 2010 Whole Year Coconut 14560.00 71300000.00 **NICOBARS** 575.00 **78** Andaman and Nicobar Islands Tapioca 69.00 2010 Whole Year 79 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Kharif Arecanut 3100.00 5200.00 80 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Other Kharif pulses 668.00 448.00 Kharif 81 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Kharif Rice 10779.00 31863.00 82 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Whole Year Banana 1531.00 12073.00 Cashewnut 83 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Whole Year 80.00 54.00 84 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 6992.00 2000 Whole Year Coconut 23900000.00 2000 Whole Year 85 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN Dry ginger 352.00 1120.00 86 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Whole Year other oilseeds 133.00 55.00 87 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Whole Year Sugarcane 141.00 3736.00 88 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Whole Year Sweet potato 843.00 78.00 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2000 Whole Year Tapioca 160.00 2108.00 90 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2001 Kharif Arecanut 3100.00 5239.00 91 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN Other Kharif pulses 2001 Kharif 418.00 199.00 92 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2001 Kharif Rice 9718.00 27033.00 93 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2001 Whole Year Cashewnut 81.00 33.00 94 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 7015.00 25250000.00 2001 Whole Year Coconut 95 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2001 Whole Year Dry ginger 344.00 1124.00 96 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2001 Whole Year Sugarcane 81.00 2379.00 97 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2001 Whole Year Sweet potato 100.00 5.00 6854.30 18995.62 98 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Kharif Rice 99 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Arecanut 1160.00 3011.70 100 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 957.57 2006 Whole Year Banana 8582.00 101 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Black pepper 175.00 191.14 102 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Cashewnut 40.50 44.00 103 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Coconut 3575.00 15400000.00 Dry chillies 104 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year 235.72 824.30 105 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Dry ginger 98.66 929.00 106 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year other oilseeds 35.62 21.40 107 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Sugarcane 71.30 1500.00 108 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Sweet potato 30.00 240.00 109 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2006 Whole Year Tapioca 255.00 1420.00 110 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 408.00 2006 Whole Year Turmeric 68.00 111 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Autumn Rice 6791.00 20118.00 112 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 73.33 889.20 Autumn Sugarcane 113 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Arecanut 1296.00 2290.00 Arhar/Tur 114 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi 294.50 90.80 115 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Black pepper 168.00 13.50 116 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Cashewnut 84.50 39.00 117 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Dry chillies 254.00 402.00 Dry ginger 118 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 110.00 1011.84 Rabi 2010 Rabi 6.80 119 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN Groundnut 4.75 120 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi 96.40 Maize 86.70 121 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Moong(Green Gram) 1168.60 556.55 122 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Sunflower 2.90 1.90 123 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Sweet potato 109.00 480.00 376.50 124 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Turmeric 68.00 125 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Rabi Urad 1059.50 458.79 126 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 657.00 7968.00 2010 Whole Year Banana 127 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Whole Year Coconut 3668.00 12700000.00 128 Andaman and Nicobar Islands NORTH AND MIDDLE ANDAMAN 2010 Whole Year 186.50 1355.00 Tapioca Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Kharif Rice 10695.80 31600.82 3105.00 5267.00 130 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Arecanut **131** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Banana 1524.00 10882.00 132 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Black pepper 487.00 102.50 133 Andaman and Nicobar Islands SOUTH ANDAMANS 2002 Whole Year Cashewnut 81.00 24.00 26830000.00 134 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Coconut 7060.00 135 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Dry chillies 239.00 20.90 136 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Dry ginger 352.00 1145.00 Andaman and Nicobar Islands SOUTH ANDAMANS 2002 Whole Year other oilseeds 40.31 455.00 **138** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2002 Whole Year Sugarcane 13760.00 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2003 Kharif Rice 10509.37 30760.70 3118.00 140 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2003 Whole Year Arecanut 5182.00 **141** Andaman and Nicobar Islands SOUTH ANDAMANS 2003 Whole Year Banana 1530.00 11558.00 142 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2003 Whole Year Black pepper 497.00 104.37 2003 Whole Year Cashewnut 143 Andaman and Nicobar Islands **SOUTH ANDAMANS** 116.50 26.14 144 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2003 Whole Year 7110.00 26660000.00 Coconut Andaman and Nicobar Islands **SOUTH ANDAMANS** 2003 Whole Year Dry chillies 340.00 578.00 2003 Whole Year 1299.20 146 Andaman and Nicobar Islands **SOUTH ANDAMANS** Dry ginger 406.00 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2003 Whole Year other oilseeds 82.09 35.51 268.00 Andaman and Nicobar Islands **SOUTH ANDAMANS** Whole Year Sugarcane 2879.52 148 2003 149 Andaman and Nicobar Islands SOUTH ANDAMANS 2004 Kharif 10681.98 29119.66 Rice Andaman and Nicobar Islands 2004 Whole Year 3160.67 3975.20 150 **SOUTH ANDAMANS** Arecanut **SOUTH ANDAMANS 151** Andaman and Nicobar Islands 2004 Whole Year Banana 1912.10 17374.40 152 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2004 Whole Year 570.00 104.36 Black pepper Andaman and Nicobar Islands **SOUTH ANDAMANS** 2004 Whole Year Cashewnut 116.50 25.10 7156.70 34750000.00 **154** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2004 Whole Year Coconut 155 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2004 Whole Year Dry chillies 340.00 822.80 410.00 156 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2004 Whole Year Dry ginger 1462.00 **157** Andaman and Nicobar Islands SOUTH ANDAMANS 2004 Whole Year other oilseeds 89.30 46.50 1560.00 158 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2004 Whole Year 145.26 Sugarcane Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Kharif Rice 8683.38 25870.98 160 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year 3250.78 3058.46 Arecanut Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Banana 1656.80 10251.13 **162** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Black pepper 656.45 35.95 163 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Cashewnut 118.50 30.22 164 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year 7050.45 26430000.00 Coconut **165** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Dry chillies 370.70 877.60 **166** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Dry ginger 200.00 1800.00 167 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year other oilseeds 53.55 34.20 168 Andaman and Nicobar Islands **SOUTH ANDAMANS** 128.02 2448.20 2005 Whole Year Sugarcane Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Sweet potato 700.35 170 Andaman and Nicobar Islands **SOUTH ANDAMANS** 1480.00 2005 Whole Year Tapioca 254.50 **171** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2005 Whole Year Turmeric 81.70 572.00 172 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 921.87 2539.64 Kharif Rice 173 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year 2000.00 2350.00 Arecanut 174 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year 483.00 4150.00 Banana 175 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year Black pepper 413.10 20.18 176 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year Cashewnut 79.50 26.00 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year Coconut 3483.09 14190000.00 178 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year Dry chillies 124.60 34.70 179 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year Dry ginger 100.90 900.00 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year 24.00 14.98 180 other oilseeds **181** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2006 Whole Year Sugarcane 53.50 999.50 **182** Andaman and Nicobar Islands **SOUTH ANDAMANS** 30.40 220.00 2006 Whole Year Sweet potato **SOUTH ANDAMANS 183** Andaman and Nicobar Islands 2006 Whole Year Tapioca 40.00 160.00 **184** Andaman and Nicobar Islands **SOUTH ANDAMANS** 60.00 2006 Whole Year Turmeric 10.00 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rice 1595.50 3788.00 Autumn **186** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Autumn Sugarcane 67.50 402.00 187 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Arecanut 1912.00 1900.00 Rabi Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Arhar/Tur 20.50 13.20 188 Rabi 189 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi Black pepper 409.00 98.00 **SOUTH ANDAMANS** 190 Andaman and Nicobar Islands 2010 15.00 10.50 Rabi Cashewnut **191** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi Dry chillies 128.00 148.00 192 Andaman and Nicobar Islands 2010 828.52 **SOUTH ANDAMANS** Rabi Dry ginger 94.00 193 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi Groundnut 15.50 7.60 194 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi Maize 253.00 73.00 **195** Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi Moong(Green Gram) 29.90 17.85 Andaman and Nicobar Islands 196 **SOUTH ANDAMANS** 2010 Rabi Sunflower 1.00 0.50 **SOUTH ANDAMANS** 2010 Rabi Sweet potato 32.00 235.00 **197** Andaman and Nicobar Islands 198 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi 10.00 105.00 Turmeric 199 Andaman and Nicobar Islands **SOUTH ANDAMANS** 2010 Rabi Urad 34.00 15.05 In [] In [65]: crop.shape (246091, 7)Out[65]: ## data set contains 246091 rows and 7 columns ## Checking info of data In [11]: crop.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 246091 entries, 0 to 246090 Data columns (total 7 columns): Column Non-Null Count 0 State_Name 246091 non-null object District_Name 246091 non-null object 1 246091 non-null datetime64[ns] Crop_Year 3 Season 246091 non-null object 4 Crop 246091 non-null object Area 5 246091 non-null float64 Production 242361 non-null float64 6 dtypes: datetime64[ns](1), float64(2), object(4) memory usage: 13.1+ MB ## 'Crop_Year' is not in proper data type, need to change to Date format In [8]: pd.to_datetime(crop['Crop_Year'], format = '%Y') In [4]: 2000-01-01 Out[4]: 2000-01-01 2 2000-01-01 3 2000-01-01 2000-01-01 246086 2014-01-01 246087 2014-01-01 246088 2014-01-01 246089 2014-01-01 246090 2014-01-01 Name: Crop_Year, Length: 246091, dtype: datetime64[ns] crop.describe() In [5]: **Production** Out[5]: Crop_Year Area count 246091.000000 2.460910e+05 2.423610e+05 mean 2005.643018 1.200282e+04 5.825034e+05 std 4.952164 5.052340e+04 1.706581e+07 min 1997.000000 4.000000e-02 0.000000e+00 25% 2002.000000 8.000000e+01 8.800000e+01 50% 2006.000000 5.820000e+02 7.290000e+02 75% 2010.000000 4.392000e+03 7.023000e+03 2015.000000 8.580100e+06 1.250800e+09 **EDA** ##Checking for missing values crop.isnull().sum() 0 State_Name Out[6]: District_Name 0 Crop_Year Season Crop Area Production 3730 dtype: int64 In [7]: ##'Production column having 3730 missing values' In [10]: crop.isnull().mean()*100 State_Name 0.000000Out[10]: District_Name 0.000000Crop_Year 0.000000 Season 0.000000 Crop 0.000000 0.000000 Area Production 1.515699 dtype: float64 In [4]: ## drop missing values of production crop.dropna(inplace = True) In [6]: crop.isnull().sum() State_Name Out[6]: District_Name 0 0 Crop_Year Season 0 Crop 0 Area 0 Production 0 dtype: int64 crop['Production'].value_counts() Production Out[9]: 4028 1.000000e+00 0.00000e+00 3523 1.000000e+02 3521 2.000000e+00 2964 3.000000e+00 2311 2.120000e+08 1 1.070000e+00 1 2.293410e+05 1 1.870600e+04 1 5.978990e+05 1 Name: count, Length: 51627, dtype: int64 crop['Crop_Year'].value_counts() In [7]: Crop_Year Out[7]: 2003 17139 2002 16536 2007 14269 2008 14230 2006 13976 2004 13858 2010 13793 2011 13791 2009 13767 2000 13553 2005 13519 2013 13475 2001 13293 2012 13184 1999 12441 1998 11262 2014 10815 1997 8899 2015 561 Name: count, dtype: int64 ## 1997 to 2015 In []: crop['State_Name'].value_counts() State_Name Out[9]: Uttar Pradesh 33306 22943 Madhya Pradesh 21122 Karnataka Bihar 18885 Assam 14628 0disha 13575 Tamil Nadu 13547 Maharashtra 12628 Rajasthan 12514 Chhattisgarh 10709 Andhra Pradesh 9628 9613 West Bengal 8436 Gujarat 5875 Haryana Telangana 5649 Uttarakhand 4896 4261 Kerala Nagaland 3906 Punjab 3173 Meghalaya 2867 Arunachal Pradesh 2546 Himachal Pradesh 2494 Jammu and Kashmir 1634 Tripura 1412 Manipur 1267 Jharkhand 1266 Mizoram 957 Puducherry 876 Sikkim 714 Dadra and Nagar Haveli 263 208 Goa Andaman and Nicobar Islands 203 Chandigarh 90 Name: count, dtype: int64 crop['District_Name'].value_counts() District_Name Out[10]: BIJAPUR 945 TUMKUR 936 **BELGAUM** 925 **HASSAN** 895 **BELLARY** 887 **HYDERABAD** 8 KHUNTI 6 RAMGARH 6 NAMSAI 1 MUMBAI Name: count, Length: 646, dtype: int64 crop['Production'].mode() 0 1.0 Out[7]: Name: Production, dtype: float64 crop['Season'].unique() ', 'Whole Year ', 'Autumn ', 'Rabi array(['Kharif Out[7]: ', 'Winter 'Summer '], dtype=object) ##Removing white spaces present in the Season columns crop['Season'] = crop['Season'].str.strip() In [6]: crop['Season'].unique() array(['Kharif', 'Whole Year', 'Autumn', 'Rabi', 'Summer', 'Winter'], dtype=object) ## Autumn & summer is same as Kharif season in India ## Winter is same as Rabi season in India ## so replacing autumn, winter with Kharif & Rabi respectively. In [8]: crop['Season'] = crop['Season'].replace({'Autumn': 'Kharif', 'Summer': 'Kharif', 'Winter': 'Rabi'}) In [9]: crop['Season'].unique() array(['Kharif', 'Whole Year', 'Rabi'], dtype=object) Out[9]: crop['State_Name'].unique() In [10]: array(['Andaman and Nicobar Islands', 'Andhra Pradesh', 'Arunachal Pradesh', 'Assam', 'Bihar', 'Chandigarh', 'Chhattisgarh', 'Dadra and Nagar Haveli', 'Goa', 'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu and Kashmir', 'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh', 'Maharashtra', 'Manipur', 'Meghalaya', 'Mizoram', 'Nagaland', 'Odisha', 'Puducherry', 'Punjab', 'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana ', 'Tripura', 'Uttar Pradesh', 'Uttarakhand', 'West Bengal'], dtype=object) In [11]: crop['State_Name'] = crop['State_Name'].str.strip() crop['Crop'] = crop['Crop'].str.strip() ## Dividing the states into North, East, South, West , Central & NorthEast for easy understanding and analysis In [14]: regions = { 'North': ['Jammu and Kashmir', 'Punjab', 'Himachal Pradesh', 'Haryana', 'Uttarakhand', 'Uttar Pradesh', 'Chandigarh'], 'East': ['Bihar', 'Odisha', 'Jharkhand', 'West Bengal'],
'South': ['Andhra Pradesh', 'Karnataka', 'Kerala', 'Tamil Nadu', 'Telangana', 'Puducherry', 'Andaman and Nicobar Islands'], 'West': ['Rajasthan' , 'Gujarat', 'Goa', 'Maharashtra', 'Dadra and Nagar Haveli'], 'Central': ['Madhya Pradesh', 'Chhattisgarh'], 'NorthEast': ['Assam', 'Sikkim', 'Nagaland', 'Meghalaya', 'Manipur', 'Mizoram', 'Tripura', 'Arunachal Pradesh'] def assign_region(state): In [15]: for region, states in regions.items(): if state in states: **return** region crop['Region'] = crop['State_Name'].apply(assign_region) crop['Region'].unique() In [17]: array(['South', 'NorthEast', 'East', 'North', 'Central', 'West'], Out[17]: dtype=object) crop['Region'].value_counts() In [18]: Region Out[18]: South 54573 North 49874 43261 East West 33397 Central 32972 NorthEast 28284 Name: count, dtype: int64 crop_df = crop.copy() In [19]: In [20]: crop_df.head() Out[20]: State_Name District_Name Crop_Year Area Production Region Season Crop **NICOBARS** Arecanut 1254.0 O Andaman and Nicobar Islands 2000 Kharif 2000.0 South 1 Andaman and Nicobar Islands **NICOBARS** 2000 Kharif Other Kharif pulses 2.0 1.0 South **NICOBARS** 2 Andaman and Nicobar Islands 2000 Kharif Rice 102.0 321.0 South 2000 Whole Year 3 Andaman and Nicobar Islands **NICOBARS** Banana 176.0 641.0 South 4 Andaman and Nicobar Islands **NICOBARS** 2000 Whole Year Cashewnut 720.0 165.0 South crop_df['Crop'].unique() In [21]: array(['Arecanut', 'Other Kharif pulses', 'Rice', 'Banana', 'Cashewnut', Out[21]: 'Coconut', 'Dry ginger', 'Sugarcane', 'Sweet potato', 'Tapioca', 'Black pepper', 'Dry chillies', 'other oilseeds', 'Turmeric', 'Maize', 'Moong(Green Gram)', 'Urad', 'Arhar/Tur', 'Groundnut', 'Sunflower', 'Bajra', 'Castor seed', 'Cotton(lint)', 'Horse-gram', 'Jowar', 'Korra', 'Ragi', 'Tobacco', 'Gram', 'Wheat', 'Masoor', 'Sesamum', 'Linseed', 'Safflower', 'Onion', 'other misc. pulses', 'Samai', 'Small millets', 'Coriander', 'Potato', 'Other Rabi pulses', 'Soyabean', 'Beans & Mutter(Vegetable)', 'Bhindi', 'Brinjal', 'Citrus Fruit', 'Cucumber', 'Grapes', 'Mango', 'Orange', 'other fibres', 'Other Fresh Fruits', 'Other Vegetables', 'Papaya', 'Pome Fruit', 'Tomato', 'Mesta', 'Cowpea(Lobia)', 'Lemon', 'Pome Granet', 'Sapota', 'Cabbage', 'Rapeseed &Mustard', 'Peas (vegetable)', 'Niger seed', 'Bottle Gourd', 'Varagu', 'Garlic', 'Ginger', 'Oilseeds total', 'Pulses total', 'Jute', 'Peas & beans (Pulses)', 'Blackgram', 'Paddy', 'Pineapple', 'Barley', 'Sannhamp', 'Khesari', 'Guar seed', 'Moth', 'Other Cereals & Millets', 'Cond-spcs other', 'Turnip', 'Carrot', 'Redish', 'Arcanut (Processed)', 'Atcanut (Raw)', 'Cashewnut Processed', 'Cashewnut Raw', 'Cardamom', 'Rubber', 'Bitter Gourd', 'Drum Stick', 'Jack Fruit', 'Snak Guard', 'Tea', 'Coffee', 'Cauliflower', 'Other Citrus Fruit', 'Water Melon', 'Total foodgrain', 'Kapas', 'Colocosia', 'Lentil', 'Bean', 'Jobster', 'Perilla', 'Rajmash Kholar', 'Ricebean (nagadal)', 'Ash Gourd', 'Beet Root', 'Lab-Lab', 'Ribed Guard', 'Yam', 'Pump Kin', 'Apple', 'Peach', 'Pear', 'Plums', 'Litchi', 'Ber', 'Other Dry Fruit', 'Jute & mesta'], dtype=object) ##creating a new column Main_cat from Crops column for easy understanding and analysis Main_cat = { In [24]: 'Cereal': ['Rice', 'Maize', 'Wheat', 'Barley', 'Varagu', 'Other Cereals & Millets', 'Ragi', 'Small millets', 'Bajra', 'Jowar', 'Paddy', 'Total foodgrain', 'Jobster'], 'Pulses': ['Moong', 'Moong(Green Gram)','Urad','Arhar/Tur','Peas & beans','Masoor','Other Kharif pulses','other misc. pulses','Ricebean (nagadal)','Rajmash 'Fruits': ['Coconut', 'Peach', 'Apple', 'Litchi', 'Pear', 'Plums', 'Ber', 'Sapota', 'Lemon', 'Pome Granet', 'Other Citrus Fruit', 'Water Melon', 'Jack Fruit', 'Grapes', 'Vegetables': ['Turnip', 'Peas', 'Beet Root', 'Carrot', 'Yam', 'Ribed Guard', 'Ash Gourd ', 'Pump Kin', 'Redish', 'Snak Guard', 'Bottle Gourd', 'Bitter Gourd', 'Cucumt 'Spices': ['Perilla', 'Ginger', 'Cardamom', 'Black pepper', 'Dry ginger', 'Garlic', 'Coriander', 'Turmeric', 'Dry chillies', 'Cond-spcs other'], 'Fibers': ['other fibres', 'Kapas', 'Jute & mesta', 'Jute', 'Mesta', 'Cotton(lint)', 'Sannhamp'], 'Oil seeds': ['Guar seed', 'Soyabean', 'other oilseeds', 'Safflower', 'Niger seed', 'Castor seed', 'Linseed', 'Sunflower', 'Rapeseed &Mustard', 'Sesamum', 'Oilseeds 'Nuts' : ['Arcanut (Processed)','Atcanut (Raw)','Cashewnut Processed','Cashewnut Raw','Cashewnut','Arecanut','Groundnut'], 'Plantcrops' : ['Tobacco', 'Coffee', 'Tea', 'Sugarcane', 'Rubber'] def assign_maincat(cat): In [25]: for cr, cats in Main_cat.items(): if cat in cats: return cr crop_df['MainCat'] = crop_df['Crop'].apply(assign_maincat) In [26]: In [29]: crop_df['MainCat'].value_counts() MainCat 63283 Cereal Pulses 55681 Oil seeds 40197 Vegetables 23154 **Spices** 21638 Nuts 11472 **Plantcrops** 10561 Fibers 9785 Fruits 6153 Name: count, dtype: int64 In [30]: ## Saving the cleaned files for further analysis and visualization in Power BI. crop_df.to_csv('crop_df.csv') crop_df.head() In [29]: Out[29]: State_Name District_Name Crop_Year Season Area Production Region MainCat Kharif Arecanut 1254.0 **0** Andaman and Nicobar Islands **NICOBARS** 2000 2000.0 South Nuts 1 Andaman and Nicobar Islands **NICOBARS** 2000 Kharif Other Kharif pulses 2.0 1.0 South Pulses 2 Andaman and Nicobar Islands **NICOBARS** 2000 Kharif 102.0 321.0 Rice South Cereal **NICOBARS 3** Andaman and Nicobar Islands 2000 Whole Year Banana 176.0 641.0 South Fruits 4 Andaman and Nicobar Islands **NICOBARS** 2000 Whole Year Cashewnut 720.0 165.0 South Nuts