



BÁO CÁO ĐỒ ÁN CUỐI KỲ

TRỰC QUAN HOÁ

CHỈ SỐ SẢN XUẤT CÔNG NGHIỆP

CỦA VIỆT NAM GIAI ĐOẠN 2017-2020

Môn học: Trực quan hoá dữ liệu

Nhóm: 25

Thành viên: 1712632 – Huỳnh Lê Minh Nhật
1712434 – Lê Thanh Hiếu

I. Thông tin nhóm:

STT	MSSV	Họ tên	Email
1	1712434	Lê Thanh Hiếu	1712434@student.hcmus.edu.vn
2	1712632	Huỳnh Lê Minh Nhật	1712632@student.hcmus.edu.vn

II. Đặt vấn đề:

- Chỉ số sản xuất công nghiệp (Index of Industrial Production - IIP) là tỷ lệ phần trăm giữa khối lượng sản xuất công nghiệp tạo ra trong kỳ hiện tại với khối lượng sản xuất công nghiệp kỳ gốc. Đây là một chỉ số quan trọng được sử dụng để phản ánh, mô tả sự phát triển của các ngành công nghiệp và phần nào đó phản ánh được những diễn biến trong nền kinh tế của Việt Nam.
- Qua thông tin báo đài, có vẻ như nền kinh tế Việt Nam đang có những dấu hiệu tích cực trong những năm gần đây (vốn đầu tư nước ngoài tăng mạnh, sự dịch chuyển các nhà máy từ Trung Quốc đến Việt Nam, một số thương hiệu Việt được ra mắt,...). Ngoài ra, trong giai đoạn từ cuối năm 2019 và 7 tháng đầu năm 2020, cả Việt Nam và thế giới đều phải hứng chịu những tổn thất rất lớn về kinh tế do đại dịch Covid-19.
- Với những lý do trên và số liệu về IIP của Cục thống kê, nhóm nhận thấy đây là một chủ đề rất thú vị để sử dụng trực quan hoá dữ liệu nhằm phân tích những nội dung sau:
 - Liệu công nghiệp Việt Nam giai đoạn 2017-2020 có phát triển đúng hướng ?
 - Đại dịch Covid-19 ảnh hưởng đến các ngành công nghiệp Việt Nam như thế nào ?

III. Tổng quan:

1. Dữ liệu:

- Dữ liệu được chọn là độ tăng trưởng chỉ số sản xuất công nghiệp của Việt Nam qua từng tháng, được thu thập từ Tổng cục thống kê.
- Dữ liệu gồm có 43 dòng (mỗi dòng ứng với 1 tháng) và 6 thuộc tính:
 - month: tháng, được thu thập từ tháng 01/2017 đến tháng 7/2020
 - iip: Chỉ số sản xuất công nghiệp toàn ngành (%)
 - mining: Khai khoáng (%)
 - manufacturing: Công nghiệp chế biến, chế tạo (%)

- energy: Sản xuất và phân phối điện, khí đốt, nước nóng, hơi nước và điều hòa không khí (%)
- water: Cung cấp nước, hoạt động quản lý và xử lý rác thải, nước thải (%)
- Các thuộc tính iip, mining, manufacturing, energy, water được lấy gốc so sánh ở năm 2015
- Nguồn: Tổng cục thống kê Việt Nam
- Link: <http://nsdp.gso.gov.vn/>

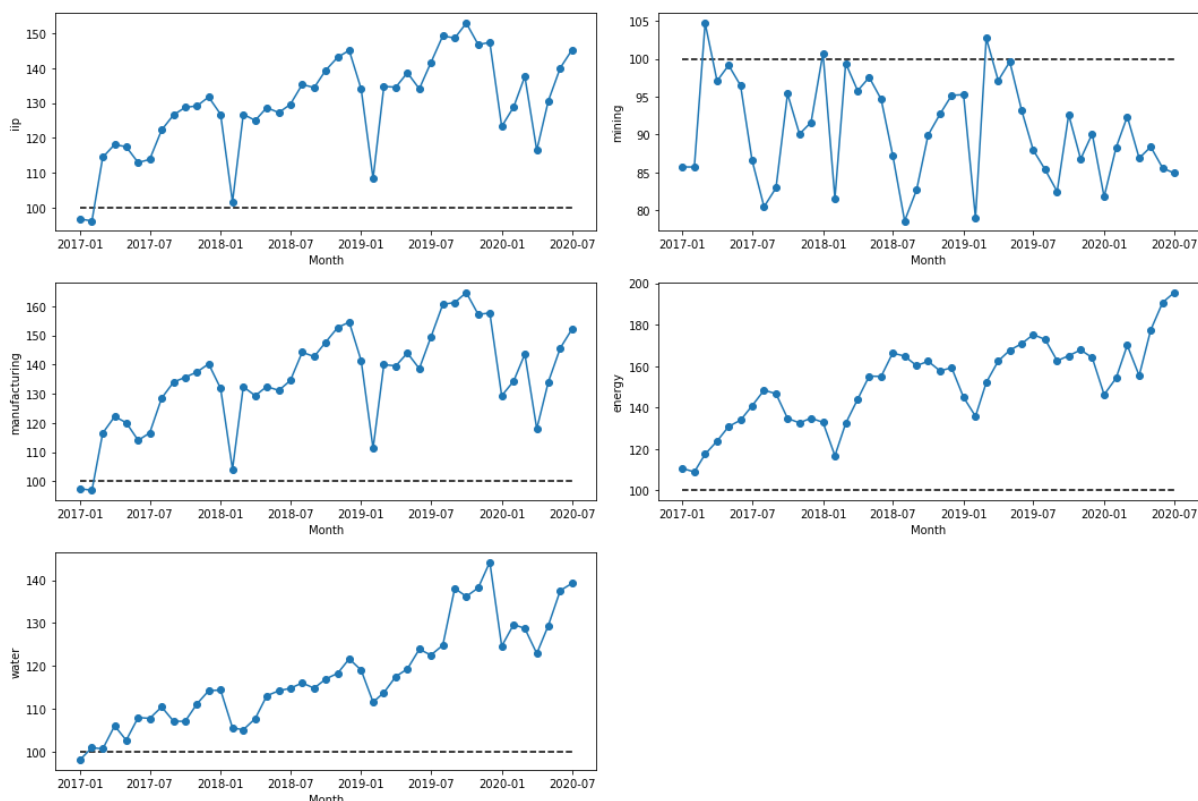
2. Thống kê:

- Thống kê cơ bản về IIP toàn ngành công nghiệp và các ngành chính

	iip	mining	manufacturing	energy	water
count	43.000000	43.000000	43.000000	43.000000	43.000000
mean	129.425349	90.510233	135.336047	151.209767	117.634419
std	13.783431	6.751565	16.511300	20.266849	11.681073
min	96.260000	78.580000	96.880000	108.890000	98.090000
25%	122.810000	85.630000	128.895000	134.850000	107.865000
50%	129.620000	90.060000	135.620000	155.030000	114.810000
75%	139.050000	95.575000	144.940000	164.930000	124.270000
max	152.900000	104.710000	164.660000	195.530000	144.260000

IV. Trục quan hoá dữ liệu :

1. Trục quan dữ liệu sử dụng biểu đồ đường (line chart)

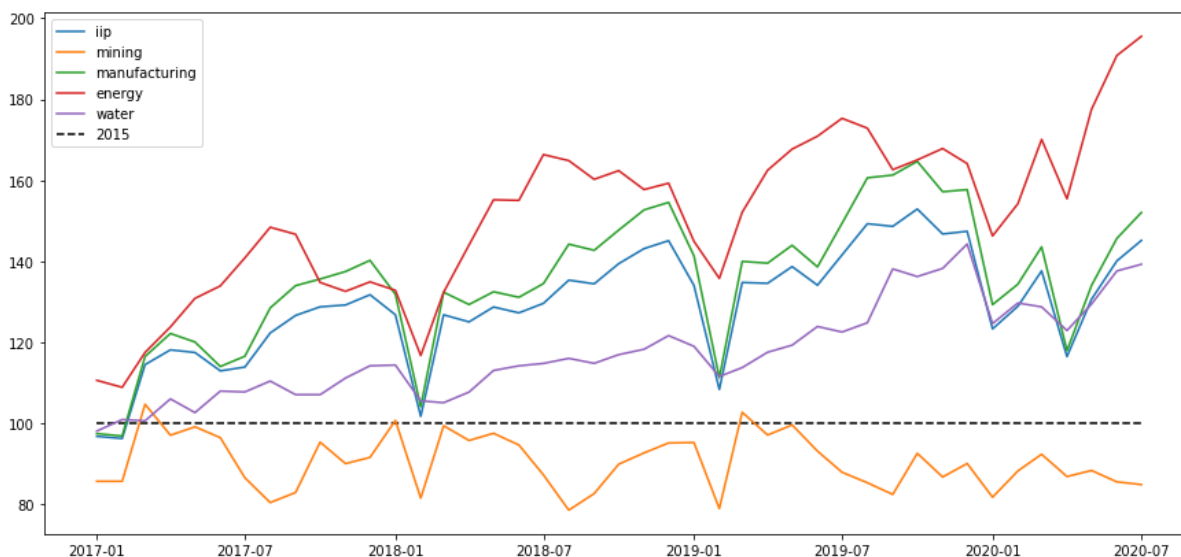


Nhận xét:

- **Chỉ số sản xuất công nghiệp - IIP**
 - Giai đoạn từ T1/2017 - T12/2019, chỉ số IIP chung của cả nước đã có những chuyển biến tích cực. Cụ thể ở mỗi tháng của những năm sau luôn có tỉ lệ tăng trưởng IIP so với năm trước đạt mức cao và so với kỳ gốc 2015.
 - Giai đoạn 7 tháng đầu năm 2020, do tình dịch Covid-19 diễn biến phức tạp trên toàn cầu làm chuỗi cung ứng nguyên liệu đầu vào bị gián đoạn, đồng thời bùng phát ở một số địa phương trên cả nước dẫn đến lệnh giãn cách xã hội đã làm suy giảm rõ rệt chỉ số IIP so với cùng kỳ năm trước.
- **Chỉ số IIP của ngành công nghiệp Sản xuất và phân phối điện, khí đốt, nước nóng, hơi nước và điều hòa không khí**
 - Giai đoạn từ T1/2017 - T12/2019, các ngành trên đều có mức tăng trưởng của mỗi tháng của những năm sau ở mức cao so với năm trước và so với kỳ gốc 2015.

- Giai đoạn 7 tháng đầu năm 2020, mặc dù tình hình dịch Covid-19 nhưng các ngành Sản xuất và phân phối điện, khí đốt, nước nóng, hơi nước và điều hòa không khí vẫn đạt mức tăng trưởng cao so với cùng kỳ 2019 (trừ tháng 4 do giãn cách xã hội nên tăng trưởng âm).
- **Chỉ số IIP của ngành Công nghiệp chế biến, chế tạo**
 - Tình hình các ngành Công nghiệp chế biến, chế tạo khá tương đương với chỉ số IIP nhưng có mức tăng trưởng cao hơn.
 - Giai đoạn từ T1/2017 - T12/2019, ở mỗi tháng của những năm sau luôn có tỉ lệ tăng trưởng so với năm trước và so với kỳ gốc 2015 đạt mức cao.
 - Giai đoạn 7 tháng đầu năm 2020, do tình dịch Covid-19 diễn biến phức tạp đã làm suy giảm rõ rệt mức tăng trưởng so với cùng kỳ năm trước, thậm chí tăng trưởng âm ở một số tháng so với cùng kỳ năm trước. Nhưng bắt đầu có dấu hiệu hồi phục.
- **Chỉ số IIP của ngành công nghiệp Khai khoáng**
 - Tình hình các ngành Khai khoáng có xu hướng giảm dần tỉ lệ tăng trưởng.
 - Giai đoạn từ T1/2017 - T12/2019, hầu như mỗi tháng của những năm sau luôn có tỉ lệ tăng trưởng thấp hơn so với cùng kỳ năm trước và so với kỳ gốc 2015.
 - Giai đoạn 7 tháng đầu năm 2020, tăng trưởng giảm sâu so với cùng kỳ.
- **Chỉ số IIP của ngành công nghiệp Cung cấp nước, hoạt động quản lý và xử lý rác thải, nước thải**
 - Tình hình các ngành Cung cấp nước, hoạt động quản lý và xử lý rác thải, nước thải khá tương đương với chỉ số IIP chung.
 - Giai đoạn từ T1/2017 - T12/2019, hầu hết mỗi tháng của những năm sau luôn có tỉ lệ tăng trưởng so với năm trước và so với kỳ gốc 2015 đạt mức cao.
 - Giai đoạn 7 tháng đầu năm 2020, do tình dịch Covid-19 diễn biến phức tạp đã làm suy giảm rõ rệt mức tăng trưởng so với cùng kỳ năm trước, thậm chí tăng trưởng âm ở một số tháng so với cùng kỳ năm trước. Nhưng bắt đầu có dấu hiệu hồi phục.

2. Trục quan hoá chỉ số IIP chung và theo từng ngành công nghiệp bằng biểu đồ đường (multiple lines chart)

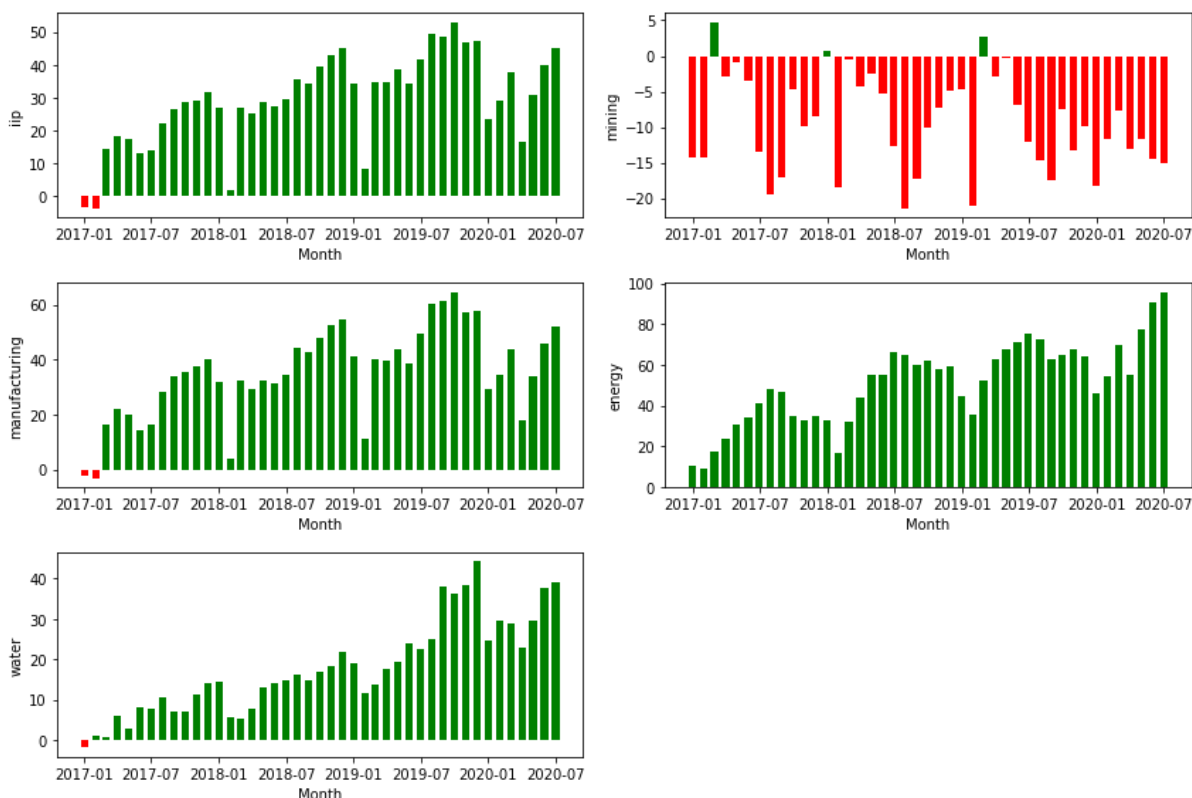


Nhận xét:

- Từ biểu đồ trên, ta có thể rút ra được một số kết luận sau:
 - Biểu đồ đường của các ngành công nghiệp có khuynh hướng giống nhau qua từng tháng trong năm. Ví dụ như giai đoạn tháng 1 - 2, tất cả các ngành công nghiệp đều có chỉ số IIP thấp nhất năm. Điều này có thể được giải thích một cách đơn giản bởi vì người lao động được nghỉ Tết vào giai đoạn này nên tình hình sản xuất công nghiệp sẽ giảm.
 - Chỉ số IIP của ngành công nghiệp khai thác và ngành công nghiệp chế biến, chế tạo có những diễn biến tích cực trong những năm gần đây. Trong đó, sự giảm dần vai trò của những ngành khai thác trong nền kinh tế để hướng đến phát triển kinh tế bền vững hơn trong tương lai là hướng đi tích cực và đúng đắn của nền kinh tế Việt Nam trong những năm gần đây. Bên cạnh đó, sự gia tăng mạnh mẽ của ngành công nghiệp chế biến, chế tạo cũng tạo ra nhiều ưu điểm tích cực hơn đối với nền kinh tế.
 - Dịch Covid-19 đã gây ra tác hại rất lớn đối với nền công nghiệp của Việt Nam, bằng chứng là nó đã làm cho chỉ số IIP của hầu hết các ngành công nghiệp (trừ Sản xuất và phân phối điện, khí đốt, nước nóng, hơi nước và điều hòa không khí) đang ở mức cao hoặc tích cực cho phát triển kinh tế (với ngành công nghiệp khai khoáng thì tăng trưởng âm dần đều là thuận lợi) trở thành tăng trưởng âm ở một số giai đoạn hoặc tăng trưởng rất thấp so với cùng kỳ năm trước. Mặc dù đã có dấu hiệu phục

hồi nhưng hầu hết tất cả các chỉ số IIP đều có mức tăng trưởng thấp nhất trong nhiều năm qua.

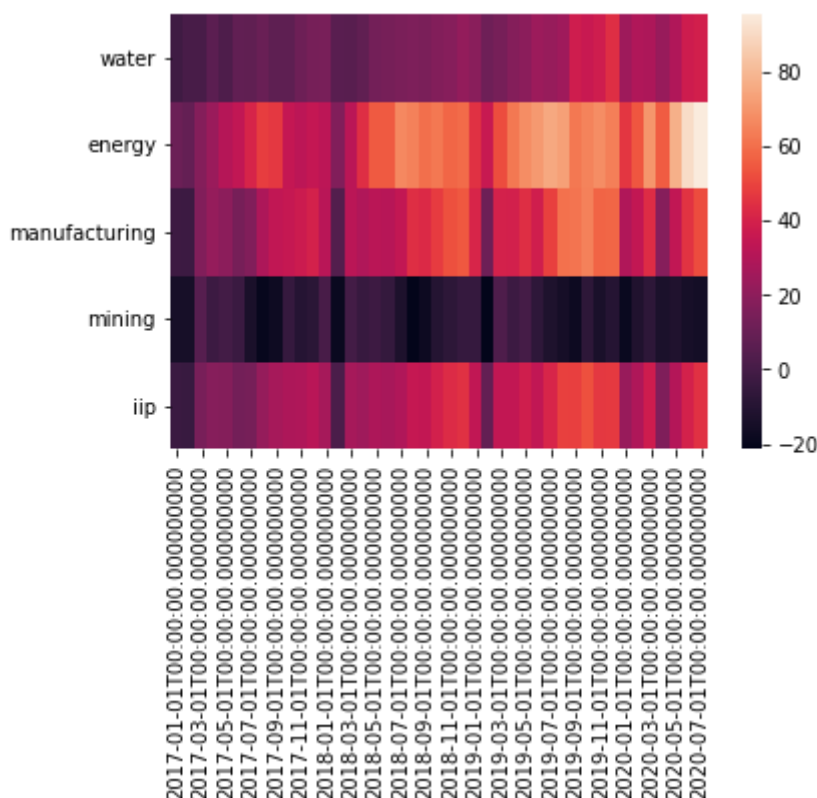
3. Trục quan dữ liệu sử dụng biểu đồ cột (bar chart)



Nhận xét:

- Nhìn chung, với biểu đồ cột, chỉ số IIP cũng được thể hiện tương tự như line chart với 2 giai đoạn T1/2017 - T12/2019 và với từng ngành công nghiệp.

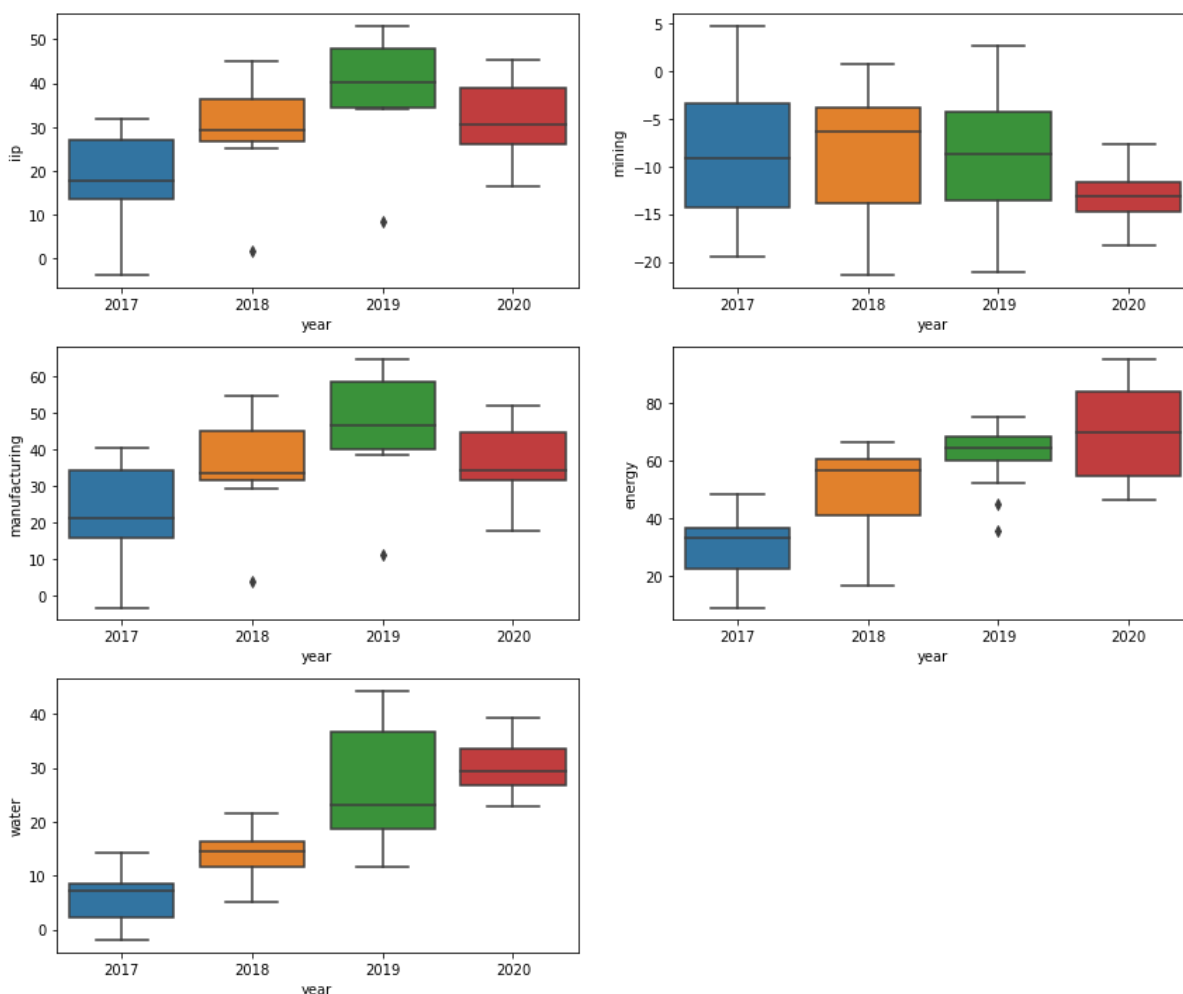
4. Trực quan dữ liệu sử dụng biểu đồ nhiệt (heatmap)



Nhận xét: Từ biểu đồ nhiệt trên, ta rút ra được một số kết luận sau:

- Giai đoạn 2017-đầu 2020: Với IIP chung, IIP manufacturing, IIP energy, IIP water thì mức độ tăng trưởng so với kỳ gốc 2015 ngày càng tăng, đặc biệt ngành energy tăng trưởng rõ rệt (bằng chứng là màu của các IIP trên chuyển từ tối đến sáng). Trong khi IIP mining duy trì ở mức rất thấp và có giảm trong giai đoạn này.
- Giai đoạn 7 tháng đầu 2020: IIP water và IIP energy có vẻ như vẫn giữ được mức tăng trưởng. Các IIP còn lại giảm so với 2019.
- Bên cạnh đó, biểu đồ nhiệt cũng cho thấy tính chu kỳ của dữ liệu khi chuyển biến của màu sắc trong từng khoảng thời gian khá tương đồng với nhau.

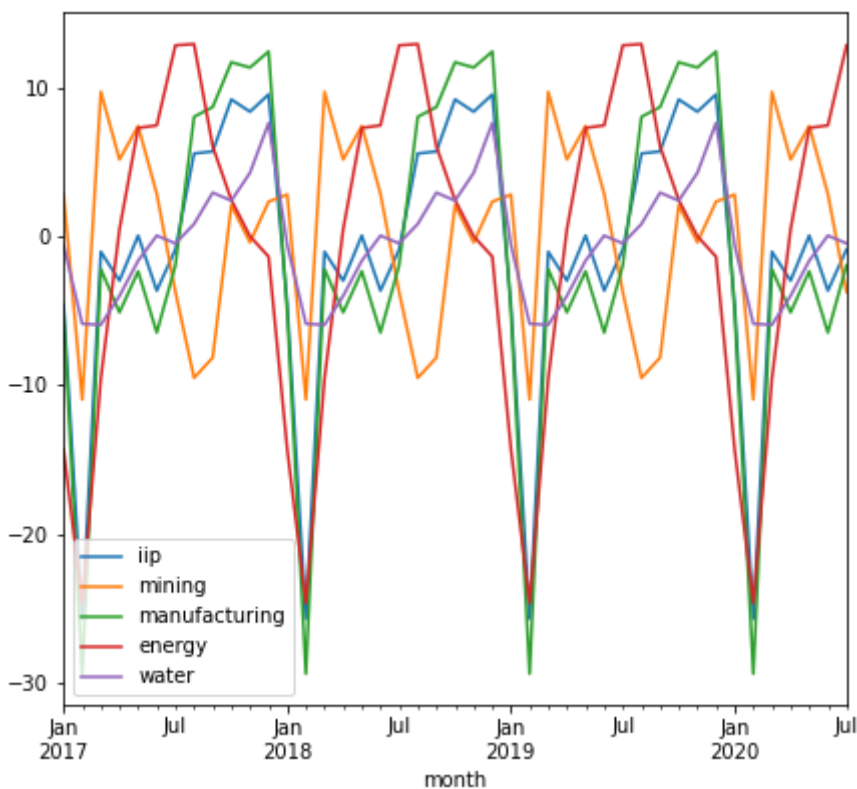
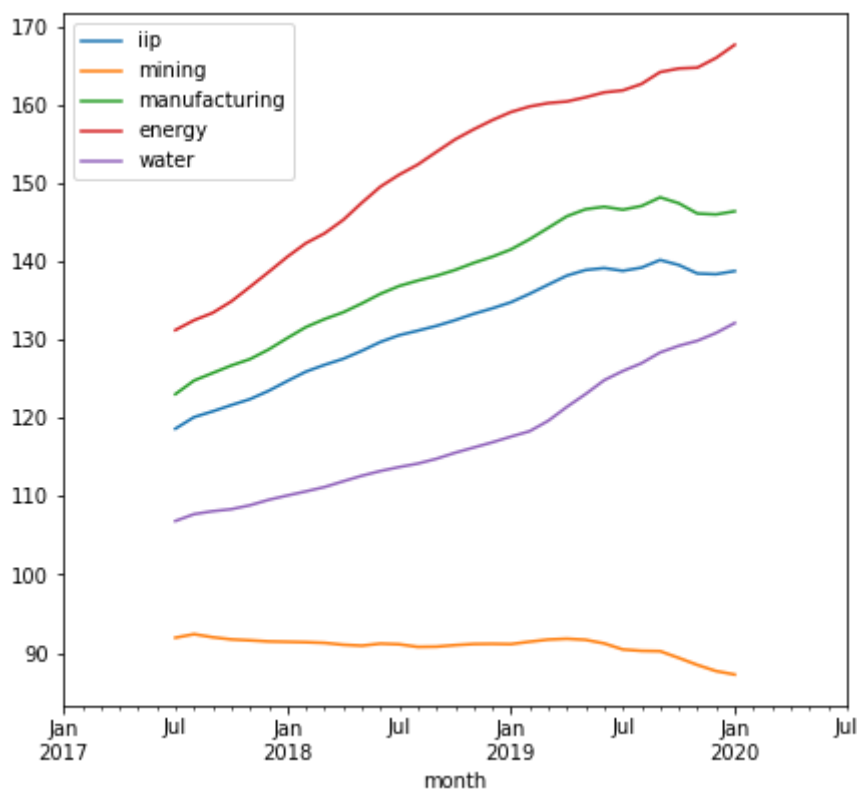
5. Trục quan dữ liệu sử dụng biểu đồ hộp (boxplot)



Nhận xét: Qua các boxplots trên, ta có được một số nhận xét sau:

- Trung vị, min, max, Q1, Q3 của các năm 2017 - 2019 tăng đều qua các năm. Điều này phản ánh sự tăng trưởng qua các năm của chỉ số IIP chung và IIP của các ngành công nghiệp nói riêng.
- Với IIP chung và IIP của ngành Công nghiệp chế biến, chế tạo và ngành xuất hiện 2 outliers vào các năm 2018, 2019. Với ngành Sản xuất và phân phối điện, khí đốt, nước nóng, hơi nước và điều hòa không khí xuất hiện 2 outliers vào năm 2019. Điều này phản ánh sự giảm sút tăng trưởng lớn của 1-2 tháng trong năm (giai đoạn Tết).
- Với IIP chung và của từng ngành công nghiệp trong giai đoạn 2020 (7 tháng đầu năm), boxplot giảm hầu hết so với năm 2019 ở các giá trị trung vị, min, max, Q1, Q3 (trừ boxplot energy).

6. Trực quan dữ liệu sử dụng biểu đồ xu hướng và biểu đồ chu kỳ (trend chart & seasonal chart)

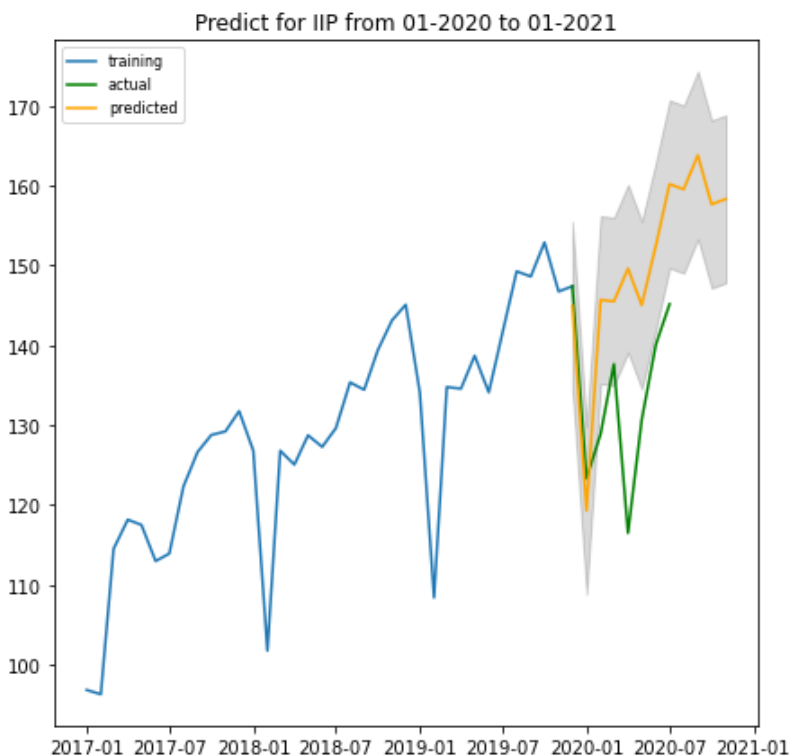


Dùng seasonal_decompose để phân tách dữ liệu thành dữ liệu dạng xu hướng và dạng chu kỳ

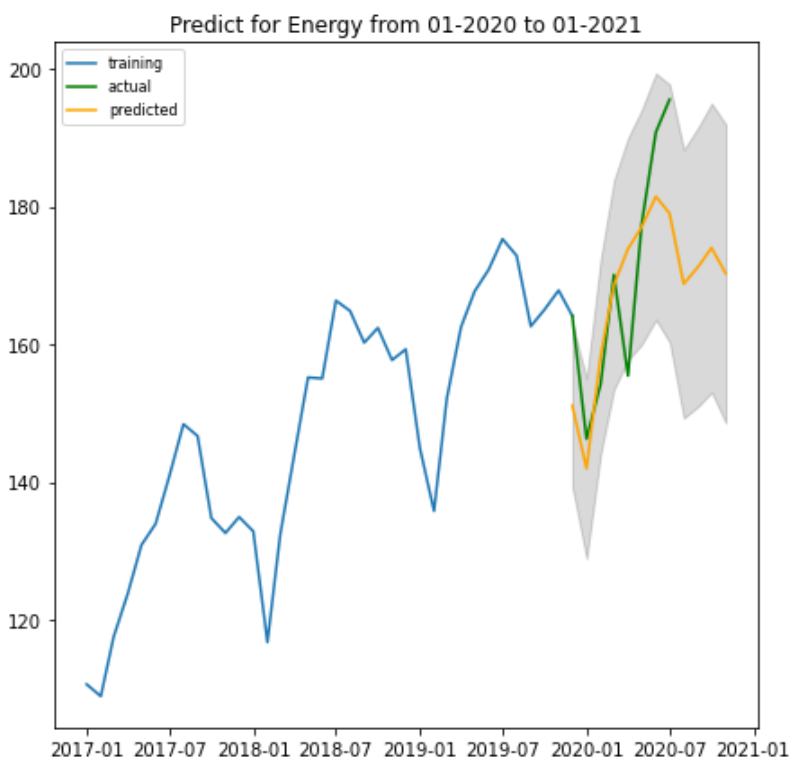
- **Ở biểu đồ xu hướng, ta có nhận xét sau:**
 - Nhìn chung, IIP chung và của từng ngành công nghiệp (trừ khai khoáng) có xu hướng tăng lên với mức độ tăng trưởng cao trong giai đoạn 2017-2020.
 - Ở giai đoạn 7 tháng đầu năm 2020 ta có IIP chung, IIP manufacturing, IIP mining giảm. Trong khi đó, IIP energy, IIP water vẫn có xu hướng tăng.
- **Ở biểu đồ chu kỳ, ta có những nhận xét sau:**
 - Đối với IIP chung và IIP từng ngành công nghiệp thì đồ thị hầu như theo một chu kỳ nhất định (cùng tăng hoặc cùng giảm ở một tháng cùng kỳ so với các tháng còn lại)

V. Áp dụng mô hình học máy:

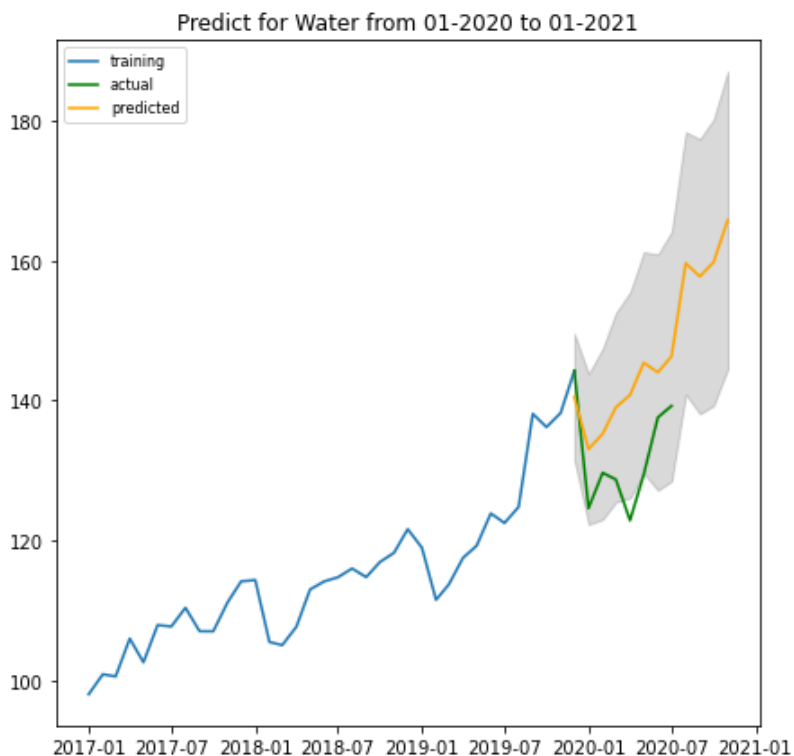
- Nhóm sử dụng mô hình SARIMA để dự đoán sự tăng trưởng của IIP từ đầu năm 2020 đến đầu năm 2021
 - Mô hình ARIMA (Autoregressive Integrated Moving Average – Tự hồi qui tích hợp Trung bình trượt): là mô hình áp dụng vào việc phân tích và dự báo dữ liệu chuỗi thời gian.
 - Mô hình SARIMA (Seasonal ARIMA): là mô hình áp dụng cho dữ liệu chuỗi thời gian có chu kỳ
- Các bước tiến hành:
 - Chia tập dữ liệu thành 2 phần:
 - Tập train: Dữ liệu từ 01/2017 đến 12/2019, dùng cho việc chọn mô hình và làm cơ sở dự đoán
 - Tập test: Dữ liệu từ 01/2020 đến 07/2020, dùng cho việc đối chiếu so sánh với kết quả dự đoán
 - Sử dụng AutoARIMA (với seasonal = True) cho việc chọn mô hình SARIMA phù hợp
 - Tiến hành dự đoán
- Kết quả dự đoán được thể hiện trong các biểu đồ bên dưới, trong đó:
 - Đường màu xanh dương là dữ liệu thực từ 01/2017 đến 12/2019
 - Đường màu xanh lá là dữ liệu thực từ 01/2020 đến 07/2020
 - Đường màu vàng là dữ liệu dự đoán từ 01/2020 đến 01/2021



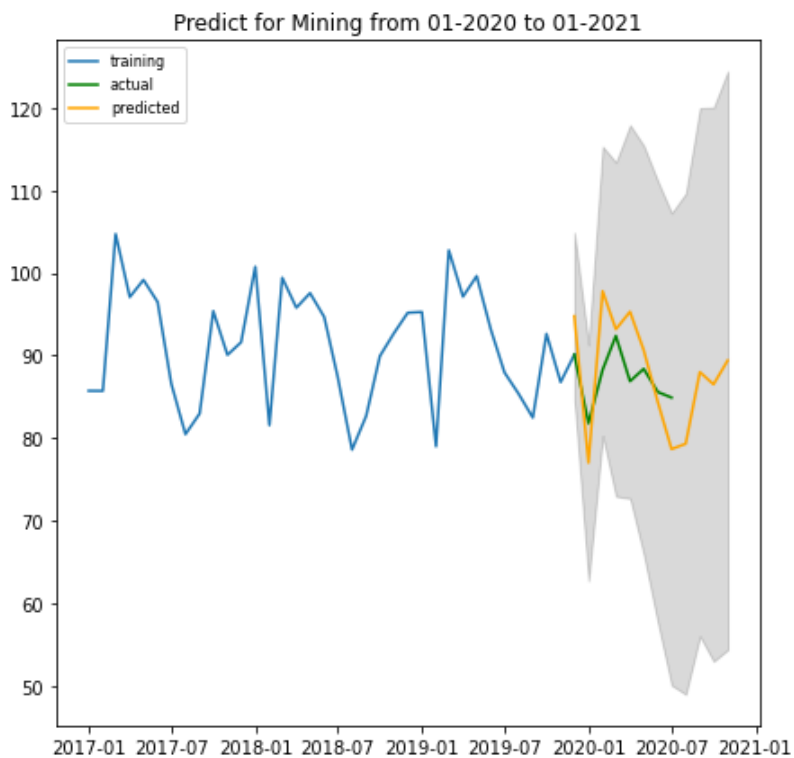
Dự đoán IIP toàn ngành từ 01-2020 đến 01-2021



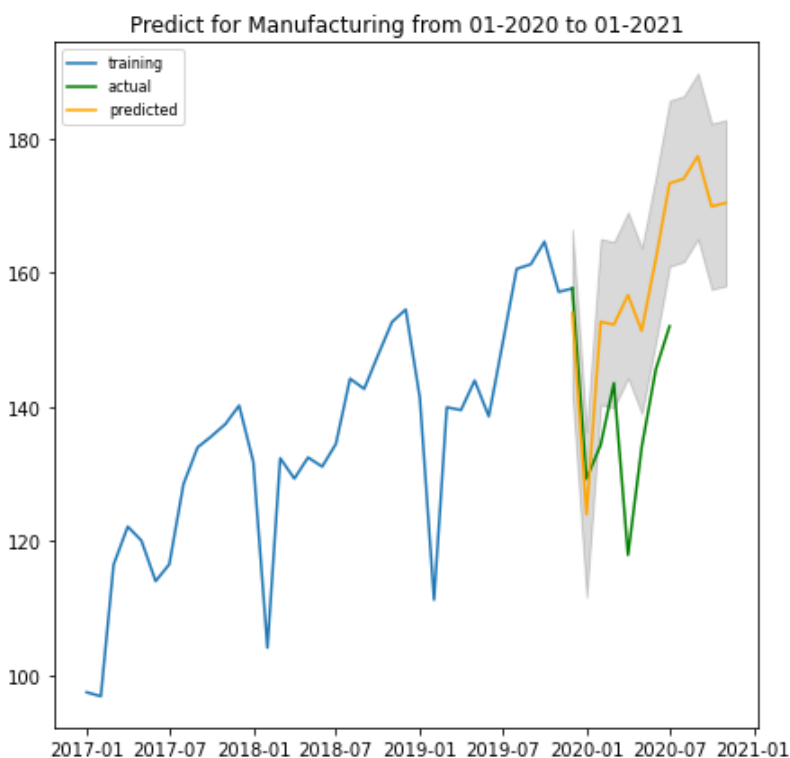
Dự đoán IIP ngành năng lượng từ 01-2020 đến 01-2021



Dự đoán IIP ngành cung cấp nước và xử lý rác thải từ 01-2020 đến 01-2021



Dự đoán IIP ngành Khai khoáng từ 01-2020 đến 01-2021



Dự đoán IIP ngành Công nghiệp chế tạo từ 01-2020 đến 01-2021

Nhận xét:

- Bằng việc áp dụng mô hình học máy dự trên dữ liệu IIP có được trong giai đoạn 2017 - cuối 2019, ta có thể dự đoán được đồ thị tăng trưởng IIP của giai đoạn đầu 2020 - đầu 2021
- Qua những biểu đồ trên, ta có thể thấy được hầu hết các giá trị dự đoán đều cao hơn so với giá trị thực tế, nghĩa là nếu như không chịu ảnh hưởng của dịch Covid-19 thì ngành công nghiệp có thể tăng trưởng rất tốt trong 7 tháng đầu năm 2020, làm tiền đề cho sự phát triển đến đầu năm 2021.

⇒ Điều này phản ánh mức độ ảnh hưởng mà dịch Covid-19 đã tác động đến chỉ số IIP cũng như ngành công nghiệp là rất lớn.

VI. Kết luận:

- Như vậy, có thể thấy chỉ số IIP là một chỉ số quan trọng trong phản ánh sức khỏe của nền kinh tế. Do đó, cần phải có những giải pháp phù hợp trong việc duy trì và thúc đẩy sự phát triển của các ngành kinh tế được thể hiện qua việc theo dõi, giám sát chặt chẽ những thay đổi của IIP.
- Từ những kết quả phân tích ở trên cho thấy chỉ số IIP của Việt Nam trong những năm qua đã có những chuyển biến tích cực theo hướng chuyển dịch cơ cấu ngành hướng đến phát triển kinh tế bền vững hơn trong tương lai.
 - Trong đó, sự giảm ngành tỷ trọng của các ngành khai thác và tăng tỷ trọng của những ngành công nghiệp chế biến, chế tạo được thể hiện rõ hơn qua chỉ số IIP qua các năm.

⇒ Điều này cho thấy sự điều hành kịp thời và hiệu quả nền kinh tế của Chính phủ cũng như nỗ lực của các bộ, ngành, địa phương trong thực hiện mục tiêu tăng trưởng. Ngoài ra, sự dịch chuyển ngành sản xuất trong nền kinh tế theo hướng phát triển bền vững là tiền đề quan trọng đối với tăng trưởng kinh tế Việt Nam trong những năm qua.

- Thông qua phân tích số liệu giai đoạn 7 tháng đầu năm 2020, qua chỉ số IIP, ta cũng thấy được hậu quả rất lớn của dịch Covid-19 đối với nền kinh tế nói chung và phát triển công nghiệp nói riêng.

⇒ Nếu như chính phủ cũng như các cá nhân, tổ chức không có những biện pháp phòng tránh dịch hiệu quả và để những đợt bùng phát dịch tiếp theo xảy ra thì nền công nghiệp sẽ phải gánh chịu những hậu quả rất nặng nề.

VII. Tham khảo:

- <http://nsdp.gso.gov.vn/>
- <https://www.gso.gov.vn/default.aspx?tabid=621>
- <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>

HẾT