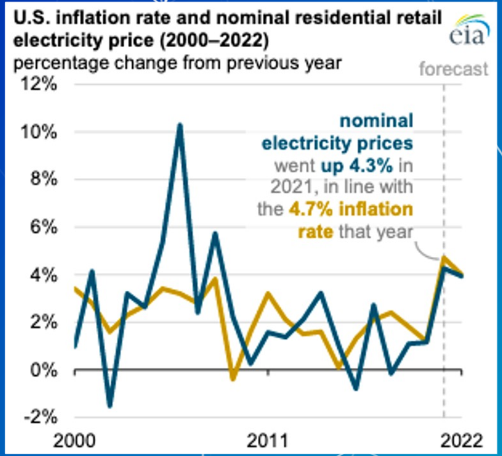
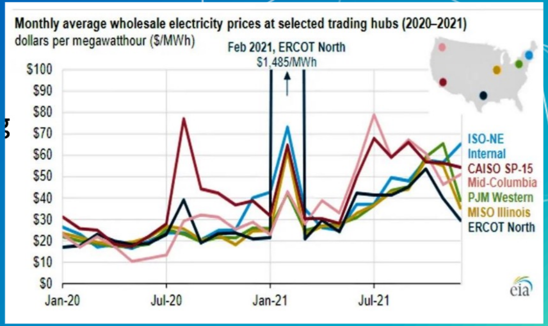
Data-Driven Electricity Trading Models and Integrated Solution Using Machine Learning Approaches

TABLE OF CONTENTS

Background

- ❑ Electricity price has been increasing consistently over the last few years
- ❑ Electricity price has become increasing volatile in recent years due to climate changes
- ❑ With the recent advancement of battery technology, there are great opportunities in trading electricity through charging and discharging a collection of batteries (Tesla PowerWall, Megapack).





Data Pre-Processing

Datasets

Smart Electricity Price Prediction project consists of four types of datasets:

- ❑ Market Price
- ❑ System Demand
- ❑ Node Location
- ❑ Weather dataset

Market Price from CAISO

©HR_SJSU

California ISO | OASIS

ATLAS REFERENCE | REPORT DEFINITION | PRICES | TRANSMISSION | SYSTEM DEMAND | ENERGY | ANCILLARY SERVICES | CONGESTION REVENUE RIGHTS | PUBLIC BIDS | RESOURCE ADEQUACY

Date From: 01/01/2021 To: 02/01/2021 Group Type SELECT_NODE Node BARRICK_LNODELON Opr Hour 01 Apply Reset

Download XML | Download CSV

Interval Locational Marginal Prices (LMP)

1 - 20 of 160

Market	Opr Date/Hour	Node	LMP Type	INTERVAL01	INTERVAL02	INTERVAL03	INTERVAL04	INTERVAL05	INTERVAL06	INTERVAL07	INTERVAL08	INTERVAL09	INTERVAL10	INTERVAL11	INTERVAL12
RTM	01/01/2021 - Hour Ending 1	BARRICK_LNODELON	LMP	25.49199	27.24280	27.95742	26.03279	26.61656	27.15095	26.43531	25.01235	24.72008	24.04886	24.04762	24.04249
RTM	01/01/2021 - Hour Ending 1	BARRICK_LNODELON	Congestion	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120
RTM	01/01/2021 - Hour Ending 1	BARRICK_LNODELON	Energy	30.27032	32.75080	33.44487	30.85066	32.13754	32.65646	31.93414	30.55384	30.27032	29.51090	29.10593	28.92930
RTM	01/01/2021 - Hour Ending 1	BARRICK_LNODELON	Loss	0.73287	0.78320	0.80375	0.75333	0.77022	0.78568	0.79237	0.74972	0.74095	0.68457	0.68454	0.68439
RTM	01/01/2021 - Hour Ending 1	BARRICK_LNODELON	Greenhouse Gas	-5.51000	-6.29000	-6.29000	-5.57000	-6.29000	-6.29000	-6.29000	-6.29000	-6.29000	-6.14541	-5.74164	-5.57000
RTM	01/02/2021 - Hour Ending 1	BARRICK_LNODELON	LMP	22.03680	22.13316	21.87821	21.71963	21.71963	21.72427	21.94539	21.93945	21.94070	21.93649	21.93524	21.88023
RTM	01/02/2021 - Hour Ending 1	BARRICK_LNODELON	Congestion	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120
RTM	01/02/2021 - Hour Ending 1	BARRICK_LNODELON	Energy	27.63848	27.73182	27.48485	27.34552	27.34552	27.35001	27.35001	27.01432	26.56552	26.56552	26.48432	26.43152
RTM	01/02/2021 - Hour Ending 1	BARRICK_LNODELON	Loss	0.68952	0.69253	0.68456	0.66532	0.66532	0.66546	0.88658	0.88634	0.88639	0.88218	0.88213	0.87991
RTM	01/02/2021 - Hour Ending 1	BARRICK_LNODELON	Greenhouse Gas	-6.29000	-6.29000	-6.29000	-6.29000	-6.29000	-6.29000	-6.29000	-5.96000	-5.51000	-5.51000	-5.43000	-5.43000
RTM	01/03/2021 - Hour Ending 1	BARRICK_LNODELON	LMP	28.75142	28.49146	26.85737	25.04212	24.89666	25.20246	26.24152	25.81738	25.48429	25.87777	24.35414	23.57459
RTM	01/03/2021 - Hour Ending 1	BARRICK_LNODELON	Congestion	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120	-0.00120
RTM	01/03/2021 - Hour Ending 1	BARRICK_LNODELON	Energy	27.65478	27.40474	25.83304	25.83304	25.83304	25.81946	27.40474	26.72015	25.83304	25.81946	25.81946	22.77199
RTM	01/03/2021 - Hour Ending 1	BARRICK_LNODELON	Loss	1.09785	1.08792	1.02552	1.00245	0.99663	1.00887	0.90208	0.88750	0.87605	0.88234	0.83039	0.80381
RTM	01/03/2021 - Hour Ending 1	BARRICK_LNODELON	Greenhouse Gas	0.00000	0.00000	0.00000	-1.79218	-1.93182	-1.62467	-2.06410	-1.78908	-1.22361	-0.82283	-2.29451	0.00000
RTM	01/04/2021 - Hour Ending 1	BARRICK_LNODELON	LMP	24.10262	28.22015	25.64638	25.09134	24.68207	24.73504	24.19345	24.14958	24.14958	23.78335	23.68677	23.59933

NTERVALSTARTTIME_GMT	INTERVALENDTIME_GMT	OPR_DT	OPR_HR	NODE_ID_XML	NODE_ID	NODE	MARKET_RUN_ID	LMP_TYPE	XML_DATA_ITEM	PNODE_RESMRID	GRP_TYPE	POS	VALUE	OPR_INTERVAL	GROUP
2020-01-01T08:00:00-00:00	2020-01-01T08:05:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	24.82750	1	5
2020-01-01T08:05:00-00:00	2020-01-01T08:10:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	26.12198	2	5
2020-01-01T08:10:00-00:00	2020-01-01T08:15:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	25.17399	3	5
2020-01-01T08:15:00-00:00	2020-01-01T08:20:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	24.87810	4	5
2020-01-01T08:20:00-00:00	2020-01-01T08:25:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	25.23021	5	5

Market Price dataset features

Features	Description
INTERVALSTARTTIME_GMT	*GMT Start Time*
INTERVALENDTIME_GMT	*GMT End Time*
OPR_DT	*Operating Date*
OPR_HR	*Operating Hour*
NODE_ID_XML	*Node Identification Code*
NODE_ID	*Node Identification Code*
NODE	*Node*
MARKET_RUN_ID	*Market identification Code*
LMP_TYPE	*Locational Marginal Price Type*
XML_DATA_ITEM	*Price Category*
PNODE_RESMRID	*Node identification Code*
GRP_TYPE	*Unused*
POS	*Planned Outage Substitution*
VALUE	*Electricity Price*
OPR_INTERVAL	*Operating Interval*
GROUP	*Group*

Demand Forecast from CAISO

©HR_SJSU

INTERVALENDTIME_GMT	OPR_DT	OPR_HR	NODE_ID_XML	NODE_ID	NODE	MARKET_RUN_ID	LMP_TYPE	...	POS	VALUE	OPR_INTERVAL	GROUP	HOUR	index	2DA	7DA	ACTUAL	DAM
2022-10-22T22:20:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	1	-0.215198	4	30	6	295371	-0.224012	-0.529867	-0.329544	-0.10695
2022-10-22T22:25:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	1	-0.286048	5	30	6	295372	-0.224012	-0.529867	-0.329544	-0.10695
2022-10-22T22:30:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	1	-0.349179	6	30	6	295373	-0.224012	-0.529867	-0.329544	-0.10695
2022-10-22T22:35:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	1	-0.172657	7	30	6	295374	-0.224012	-0.529867	-0.329544	-0.10695
2022-10-22T22:40:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	1	-0.125584	8	30	6	295375	-0.224012	-0.529867	-0.329544	-0.10695

Demand Forecast dataset features

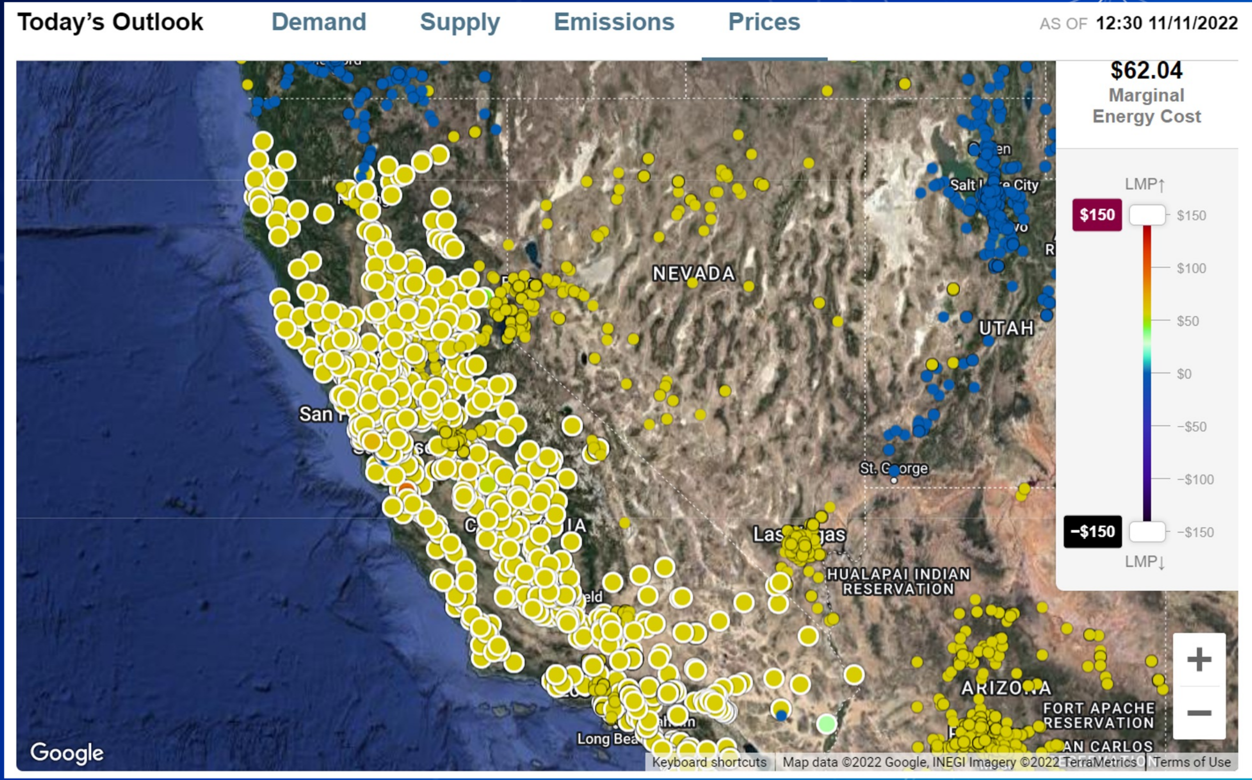
Features	Description
INTERVALSTARTTIME_GMT	*GMT Start Time*
INTERVALENDTIME_GMT	*GMT End Time*
LOAD_TYPE	*Load type*
OPR_DT	*Operating Date*
OPR_HR	*Operating Hour*
OPR_INTERVAL	*Operating Interval*
MARKET_RUN_ID	*Market identification Code*
TAC_AREA_NAME	*Transmission Access Charge Area Name*
XML_DATA_ITEM	*Price Category*
POS	*Planned Outage Substitution*
MW	*Mega Watt*
2DA	*2 days ahead forecast*
7DA	*7 days ahead forecast*
Actual	*Actual demand*
DAM	*Day ahead Market*

Data Collection

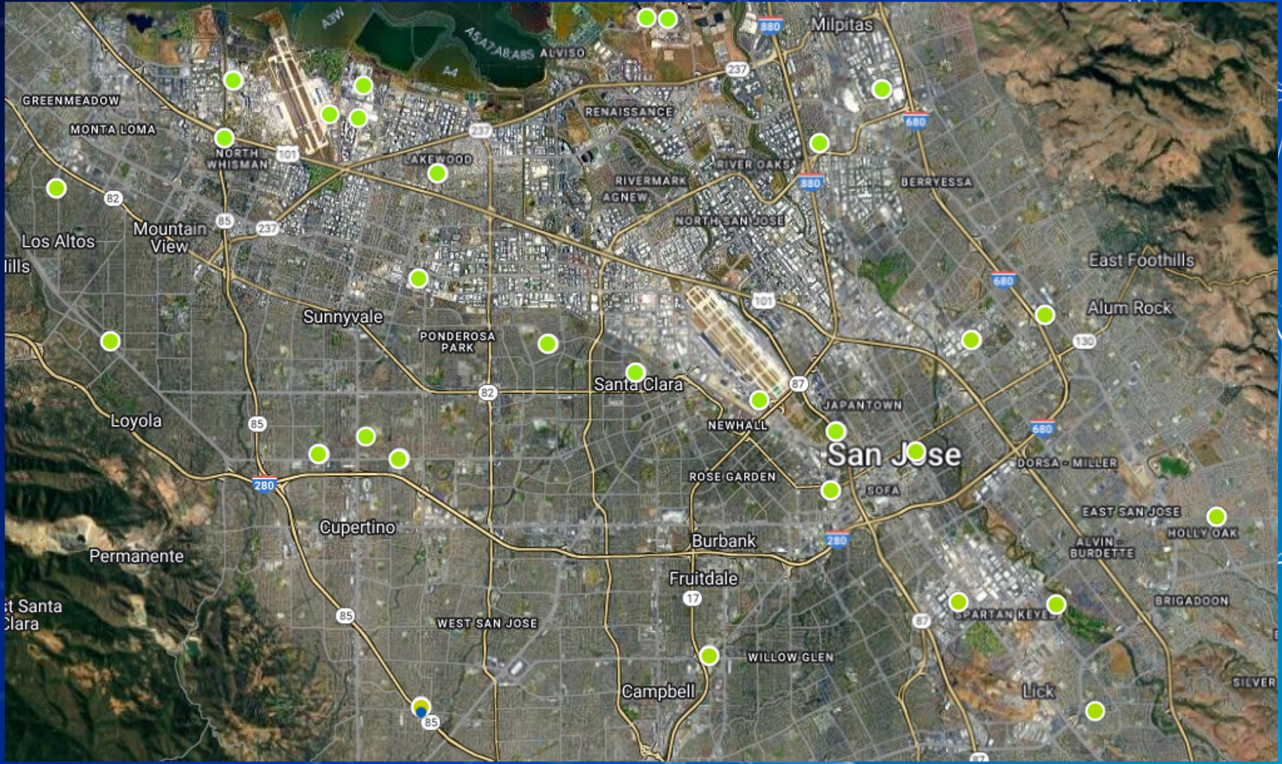
- Python script to collect data
- Acknowledgement Dr. Wang

```
Here we only get the price data from one node, since marginal cost of energy is the same throughout the whole network.
Differences in locational marignal prices are based on things like congestion, line losses, GHG, etc
RTPD = real-time pre-dispatch market, a.k.a. fifteen-minute market
'''
#20150101   20180101
def get_rtm_prices(node, start_date = '20220917', end_date = '20220918', market_run_id='RTM'):
    base_url = "http://oasis.caiso.com/oasisapi/SingleZip?"

    # Add timedelta of 8 hrs to account for timezone (times for data requests are in GMT)
    start_dt = datetime.strptime(start_date, '%Y%m%d') + timedelta(hours=8)
    start = start_dt

    end_dt = datetime.strptime(end_date, '%Y%m%d') + timedelta(hours=8)

    # Download 30 days of data at a time. The max # of days per request is 31.
    end = min([end_dt, start_dt + timedelta(days=30)])
    csv_file = "RT_market_data_" + start_date + "_" + end_date + "_" + node + '.csv'

    #print(csv_file)

    while (start_dt < end_dt):
        # get url query for data in this timeframe; read it to file; extract contents from file; rewrite contents into a joint file
        url = base_url + 'resultformat=6&queryname=PRC_INTVL_LMP&version=3&startdatetime=' + start_dt.strftime('%Y%m%dT%H:%M-0000') + '&enddatetime=' + end.st
        #print(url)
        urllib.request.urlretrieve(url, 'temp.zip')
        tempzip = ZipFile('temp.zip')
        filename = tempzip.namelist()[0]
        data_string = tempzip.read(filename).decode("utf-8")
        fmm_data = pd.read_csv(StringIO(data_string))
        #put heaters into the file only if it's the first line of the file, so we don't have extraneous random headers
        fmm_data.to_csv(csv_file, mode='a', index=False, header=(start == start_dt))

        start_dt = end
        end = min([end_dt, start_dt + timedelta(days=30)])
```

Node Location

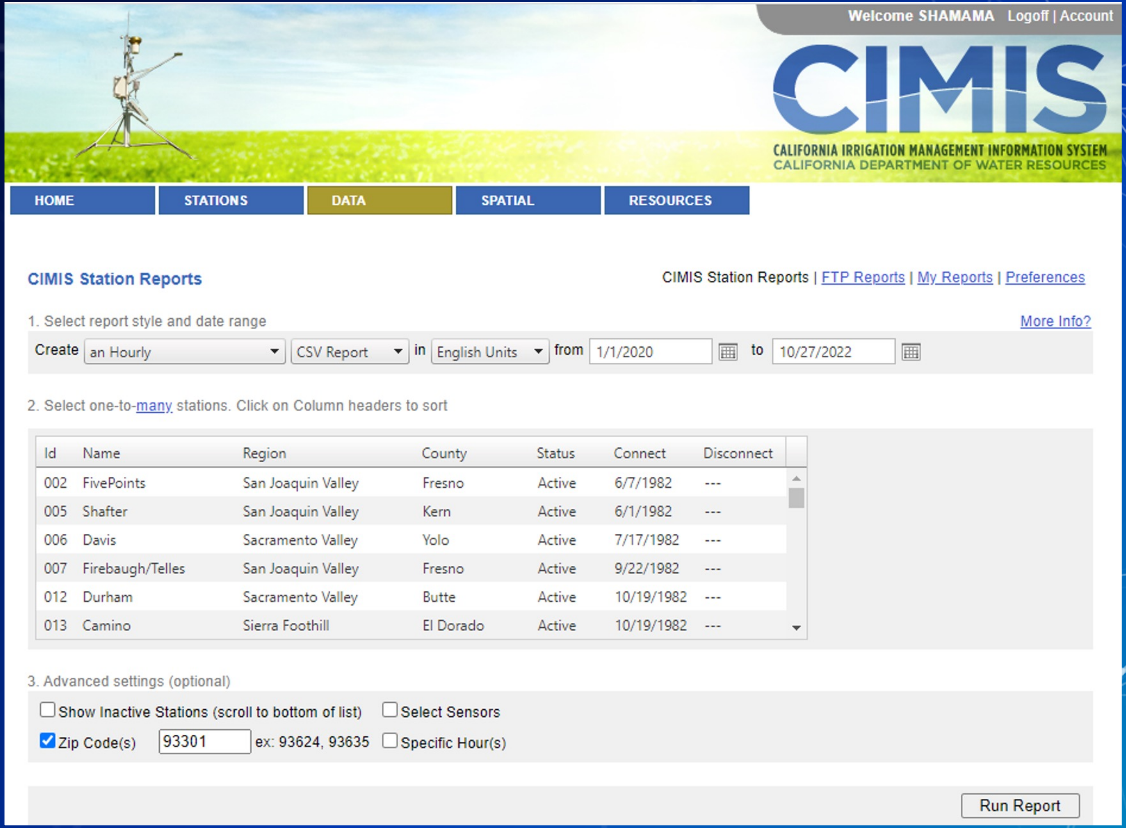


10 PNodes in the Bay Area



SNJSEA_1_N101
SNJOSEB_1_N013
CATALYST_7_N002
FMC_1_N003
CSCGNRA2_7_N101
EVRGREEN_1_N019
MONTAGUE_1_N007
MILPITAS_1_N008
DIXONLD_1_N008
LAWRENCE_1_N006

Weather Dataset



Features	Description
CIMIS Region	*Location*
Date	*Date*
Hour (PST)	*Hour*
ETo (in)	*Evapotranspiration*
Precip (in)	*Precipitation*
Rel Hum (%)	*Relative Humidity*
Dew Point (F)	*Dew Point*
Wind Speed (mph)	*Wind Speed*
Wind Dir (0-360)	*Wid Direction*
Soil Temp (F)	*Soil Temperature*

Data Preprocessing

- ❑ Filtered to include only real-time market data = LMP in the Market Price dataset
- ❑ Filtered by TAC Area: Transmission access charge (TAC) area in Demand dataset
- ❑ Interval Start Time and End Time in GMT. This time needs to be converted to California local time
- ❑ Handling Variability: Price normalization
- ❑ Joining 3 Datasets on time

Samples

©HR_SJSU

Real time market data

	INTERVALSTARTTIME_GMT	INTERVALENDTIME_GMT	OPR_DT	OPR_HR	NODE_ID_XML	NODE_ID	NODE	MARKET_RUN_ID	LMP_TYPE	XML_DATA_ITEM	PNODE_RESMRID	GRP_TYPE	POS	VALUE	OPR_INTERVAL	GROUP
1152	2020-01-01T08:00:00-00:00	2020-01-01T08:05:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	-0.291858	1	5
1153	2020-01-01T08:05:00-00:00	2020-01-01T08:10:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	-0.270029	2	5
1154	2020-01-01T08:10:00-00:00	2020-01-01T08:15:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	-0.286015	3	5
1155	2020-01-01T08:15:00-00:00	2020-01-01T08:20:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	-0.291005	4	5
1156	2020-01-01T08:20:00-00:00	2020-01-01T08:25:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	0096WD_7_N001	ALL	1	-0.285067	5	5

Demand forecast data

	INTERVALSTARTTIME_GMT	INTERVALENDTIME_GMT	OPR_DT	OPR_HR	NODE_ID_XML	NODE_ID	NODE	MARKET_RUN_ID	LMP_TYPE	XML_DATA_ITEM	...	HOUR	2DA	7DA	ACTUAL	DAM	ETo (in)	Precip (in)	Rel Hum (%)	Wind Speed (mph)	Soil Temp (F)
295366	2022-10-22T21:50:00-00:00	2022-10-22T21:55:00-00:00	2022-10-22	15	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	5	-0.181786	-0.492943	-0.415387	-0.061958	-0.691962	-0.07183	1.123573	-1.029871	0.293208
295367	2022-10-22T21:55:00-00:00	2022-10-22T22:00:00-00:00	2022-10-22	15	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	5	-0.181786	-0.492943	-0.415387	-0.061958	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295368	2022-10-22T22:00:00-00:00	2022-10-22T22:05:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295369	2022-10-22T22:05:00-00:00	2022-10-22T22:10:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295370	2022-10-22T22:10:00-00:00	2022-10-22T22:15:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295371	2022-10-22T22:15:00-00:00	2022-10-22T22:20:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295372	2022-10-22T22:20:00-00:00	2022-10-22T22:25:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295373	2022-10-22T22:25:00-00:00	2022-10-22T22:30:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295374	2022-10-22T22:30:00-00:00	2022-10-22T22:35:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015
295375	2022-10-22T22:35:00-00:00	2022-10-22T22:40:00-00:00	2022-10-22	16	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	LMP_PRC	...	6	-0.223785	-0.476465	-0.389439	-0.107000	-0.691962	-0.07183	1.164881	-1.029871	0.277015

Weather data

	INTERVALSTARTTIME_GMT	INTERVALENDTIME_GMT	LOAD_TYPE	OPR_DT	OPR_HR	OPR_INTERVAL	MARKET_RUN_ID	TAC_AREA_NAME	LABEL	XML_DATA_ITEM	POS	MW	EXECUTION_TYPE	GROUP
201	2020-01-01T08:00:00-00:00	2020-01-01T09:00:00-00:00	2	2020-01-01	1	0	2DA	PGE	Demand Forecast 2-Day Ahead	SYS_FCST_2DA_MW	1.8	-0.934468	2DA	9
547	2020-01-01T08:00:00-00:00	2020-01-01T09:00:00-00:00	0	2020-01-01	1	0	7DA	PGE	Demand Forecast 7-Day Ahead	SYS_FCST_7DA_MW	0.8	-0.946311	7DA	23
908	2020-01-01T08:00:00-00:00	2020-01-01T09:00:00-00:00	0	2020-01-01	1	0	ACTUAL	PGE	Total Actual Hourly Integrated Load	SYS_FCST_ACT_MW	3.8	-0.998405	ACTUAL	38
1232	2020-01-01T08:00:00-00:00	2020-01-01T09:00:00-00:00	1	2020-01-01	1	0	DAM	PGE	Demand Forecast Day Ahead	SYS_FCST_DA_MW	2.8	-0.993347	DAM	52
196	2020-01-01T09:00:00-00:00	2020-01-01T10:00:00-00:00	2	2020-01-01	2	0	2DA	PGE	Demand Forecast 2-Day Ahead	SYS_FCST_2DA_MW	1.8	-1.116787	2DA	9
528	2020-01-01T09:00:00-00:00	2020-01-01T10:00:00-00:00	0	2020-01-01	2	0	7DA	PGE	Demand Forecast 7-Day Ahead	SYS_FCST_7DA_MW	0.8	-1.108707	7DA	23
911	2020-01-01T09:00:00-00:00	2020-01-01T10:00:00-00:00	0	2020-01-01	2	0	ACTUAL	PGE	Total Actual Hourly Integrated Load	SYS_FCST_ACT_MW	3.8	-1.173412	ACTUAL	38
1236	2020-01-01T09:00:00-00:00	2020-01-01T10:00:00-00:00	1	2020-01-01	2	0	DAM	PGE	Demand Forecast Day Ahead	SYS_FCST_DA_MW	2.8	-1.178878	DAM	52
215	2020-01-01T10:00:00-00:00	2020-01-01T11:00:00-00:00	2	2020-01-01	3	0	2DA	PGE	Demand Forecast 2-Day Ahead	SYS_FCST_2DA_MW	1.8	-1.249049	2DA	9
538	2020-01-01T10:00:00-00:00	2020-01-01T11:00:00-00:00	0	2020-01-01	3	0	7DA	PGE	Demand Forecast 7-Day Ahead	SYS_FCST_7DA_MW	0.8	-1.235959	7DA	23

Final Joined Sample

- ❑ Sample of the final dataset

```
df.head()
```

RVALSTARTTIME_GMT	Unnamed: 0	index	INTERVALENDTIME_GMT	OPR_DT	OPR_HR	NODE_ID_XML	NODE_ID	NODE	MARKET_RUN_ID	LMP_TYPE	...	2DA	7DA	ACTUAL	DAM	index.2	ETo (in)	Precip (in)	R
20-01-01T08:00:00-00:00	0	-1.730677	2020-01-01T08:05:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	2020.67	-0.910149	-1.047574	-0.983051	-1.731989	-0.691962	-0.07183	0.(
20-01-01T08:05:00-00:00	1	-1.730675	2020-01-01T08:10:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	2020.67	-0.910149	-1.047574	-0.983051	-1.731989	-0.691962	-0.07183	0.(
20-01-01T08:10:00-00:00	2	-1.730673	2020-01-01T08:15:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	2020.67	-0.910149	-1.047574	-0.983051	-1.731989	-0.691962	-0.07183	0.(
20-01-01T08:15:00-00:00	3	-1.730670	2020-01-01T08:20:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	2020.67	-0.910149	-1.047574	-0.983051	-1.731989	-0.691962	-0.07183	0.(
20-01-01T08:20:00-00:00	4	-1.730668	2020-01-01T08:25:00-00:00	2020-01-01	1	0096WD_7_N001	0096WD_7_N001	0096WD_7_N001	RTM	LMP	...	2020.67	-0.910149	-1.047574	-0.983051	-1.731989	-0.691962	-0.07183	0.(

s × 29 columns

Feature Engineering

- Input features: 'INTERVALSTARTTIME_GMT', ‘INTERVALENDTIME_GMT', 'OPR_DT', 'OPR_HR', 'NODE_ID_XML', 'NODE_ID', 'NODE', 'MARKET_RUN_ID', 'LMP_TYPE', 'XML_DATA_ITEM', 'PNODE_RESMRID', 'GRP_TYPE', 'OPR_INTERVAL', 'GROUP', 'HOUR', '2DA', '7DA', 'ACTUAL', 'DAM'', 'ETo (in)', 'Precip (in)', 'Rel Hum (%)', 'Wind Speed (mph)', 'Soil Temp (F)'.
- Output feature: ‘VALUE’ which represents the electricity price
- Categorical features: "INTERVALENDTIME_GMT", "OPR_DT", "OPR_HR", "NODE_ID_XML", "NODE_ID", "NODE", MARKET_RUN_ID", "LMP_TYPE", "XML_DATA_ITEM","PNODE_RESMRID", "GRP_TYPE"
- Set Index: 'INTERVALSTARTTIME_GMT'
- Final 13 features: 'POS', 'OPR_INTERVAL', 'GROUP', 'HOUR', '2DA', '7DA', 'ACTUAL', 'DAM', 'ETo (in)', 'Precip (in)', 'Rel Hum (%)', 'Wind Speed (mph)',
- 'Soil Temp (F)'

Training and Test Data Preparation

Features in Train/Test Dataset

Weather	Demand	Market Data
• 'ETo (in)' • 'Precip (in)’ • 'Rel Hum (%)' • 'Wind Speed (mph)' • “'Soil Temp (F)'	• '2DA' • '7DA' • 'ACTUAL' • 'DAM'	• 'POS' • 'OPR_INTERVAL’ • 'GROUP' • HOUR **Value is used for evaluation metrics**

	POS	VALUE	OPR_INTERVAL	GROUP	HOUR	2DA	7DA	ACTUAL	DAM	ETo (in)	Precip (in)	Rel Hum (%)	Wind Speed (mph)	Soil Temp (F)
INTERVALSTARTTIME_GMT														
2020-01-01 08:00:00+00:00	1	-0.422382	1	5	16	-4.146099	-4.113026	-4.485045	-4.307317	-3.023405	-0.03305	3.663296	-1.213782	0.188814
2020-01-01 08:05:00+00:00	1	-0.405158	2	5	16	-4.146099	-4.113026	-4.485045	-4.307317	-3.023405	-0.03305	3.663296	-1.213782	0.188814
2020-01-01 08:10:00+00:00	1	-0.417772	3	5	16	-4.146099	-4.113026	-4.485045	-4.307317	-3.023405	-0.03305	3.663296	-1.213782	0.188814
2020-01-01 08:15:00+00:00	1	-0.418884	4	5	16	-4.146099	-4.113026	-4.485045	-4.307317	-3.023405	-0.03305	3.663296	-1.213782	0.188814
2020-01-01 08:20:00+00:00	1	-0.414159	5	5	16	-4.146099	-4.113026	-4.485045	-4.307317	-3.023405	-0.03305	3.663296	-1.213782	0.188814

Train and Test Data Comparison

	Raw	Processed	Train	Test
Time Interval	2020-2022	2020-2022	2020-2021	2022
# of observations	3,125,952	3,125,952	3,047,328	78,624
# of attributes	28	13	13	13

Train and Test Data Comparison

©HR_SJSU

- Train data

	POS	VALUE	OPR_INTERVAL	GROUP	HOUR	2DA	7DA	ACTUAL	DAM	ETo (in)
count	3047328.0	3.047328e+06	3.047328e+06	3.047328e+06	3.047328e+06	3.047328e+06	3.047328e+06	3.047328e+06	3.047328e+06	3.047328e+06
mean	1.0	-2.844799e-02	6.500000e+00	7.715886e+01	1.150000e+01	-3.698669e-03	-5.637792e-03	-4.497926e-03	-5.758065e-03	3.438575e-03
std	0.0	9.910099e-01	3.452053e+00	4.327156e+01	6.923061e+00	1.007425e+00	1.006209e+00	1.006839e+00	1.005638e+00	1.001733e+00
min	1.0	-1.610701e+01	1.000000e+00	5.000000e+00	0.000000e+00	-6.190183e+00	-7.290897e+00	-7.068154e+00	-7.236638e+00	-3.023405e+00
25%	1.0	-4.027381e-01	3.750000e+00	4.000000e+01	5.000000e+00	2.813929e-01	1.878294e-01	1.942408e-01	1.833645e-01	1.629812e-01
50%	1.0	-2.169648e-01	6.500000e+00	7.500000e+01	1.150000e+01	2.813929e-01	1.878294e-01	1.942408e-01	1.833645e-01	1.629812e-01
75%	1.0	8.489525e-02	9.250000e+00	1.150000e+02	1.800000e+01	2.813929e-01	1.878294e-01	1.942408e-01	1.833645e-01	1.629812e-01
max	1.0	4.933787e+01	1.200000e+01	1.550000e+02	2.300000e+01	6.710989e+00	1.114840e+01	1.021789e+01	1.212198e+01	6.535753e+00

- Test data

	POS	VALUE	OPR_INTERVAL	GROUP	HOUR	2DA	7DA	ACTUAL	DAM	ETo (in)	Precip (i
count	78624.0	78624.000000	78624.000000	78624.000000	78624.00000	78624.000000	78624.000000	78624.000000	78624.000000	78624.000000	78624.0000
mean	1.0	1.102594	6.500000	77.527473	11.50000	0.143354	0.218511	0.174332	0.223172	-0.132176	0.1775
std	0.0	0.668336	3.452074	43.039826	6.88828	0.633402	0.684292	0.660982	0.714588	0.923785	2.5066
min	1.0	-0.752880	1.000000	5.000000	0.00000	-4.061572	-3.756081	-4.200013	-3.972005	-3.023405	-0.0330
25%	1.0	0.810305	3.750000	40.000000	6.00000	0.281393	0.187829	0.194241	0.183365	0.162981	-0.0330
50%	1.0	1.092698	6.500000	80.000000	11.50000	0.281393	0.187829	0.194241	0.183365	0.162981	-0.0330
75%	1.0	1.384106	9.250000	115.000000	17.00000	0.281393	0.187829	0.194241	0.183365	0.162981	-0.0330
max	1.0	12.976391	12.000000	150.000000	23.00000	4.388507	5.291954	5.514850	6.057222	0.162981	68.3733

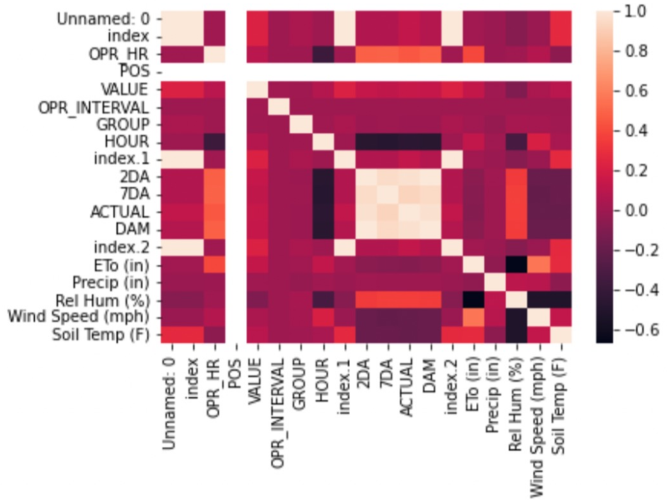
Data Analytics Results

Data Statistics

- VALUE is our variable of interest
- Correlation heat map shows high correlation between the demand variables (2DA, 7DA, ACTUAL, DAM), operating hour, and relative humidity

```
sns.heatmap(df.corr())
```

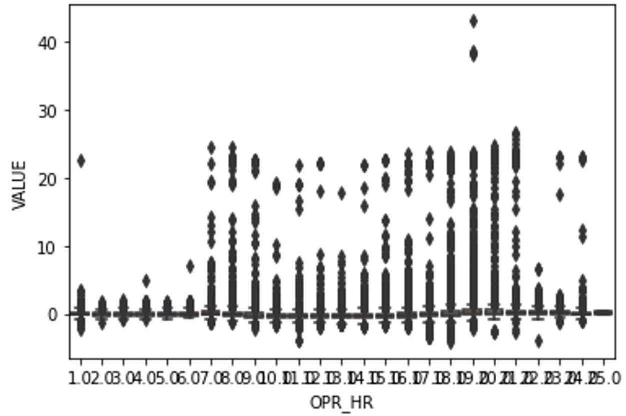
<AxesSubplot:>

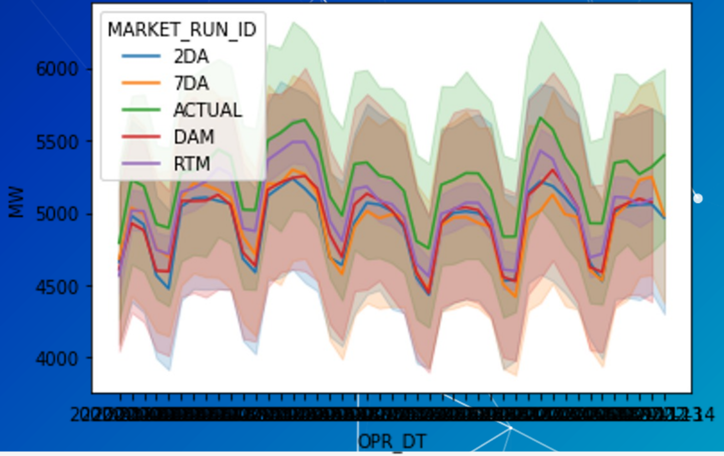


Data Statistics

- normalized data has values between 0-1.

```
sns.boxplot(x='OPR_HR', y='VALUE', data=df, palette='Accent')
plt.show()
```

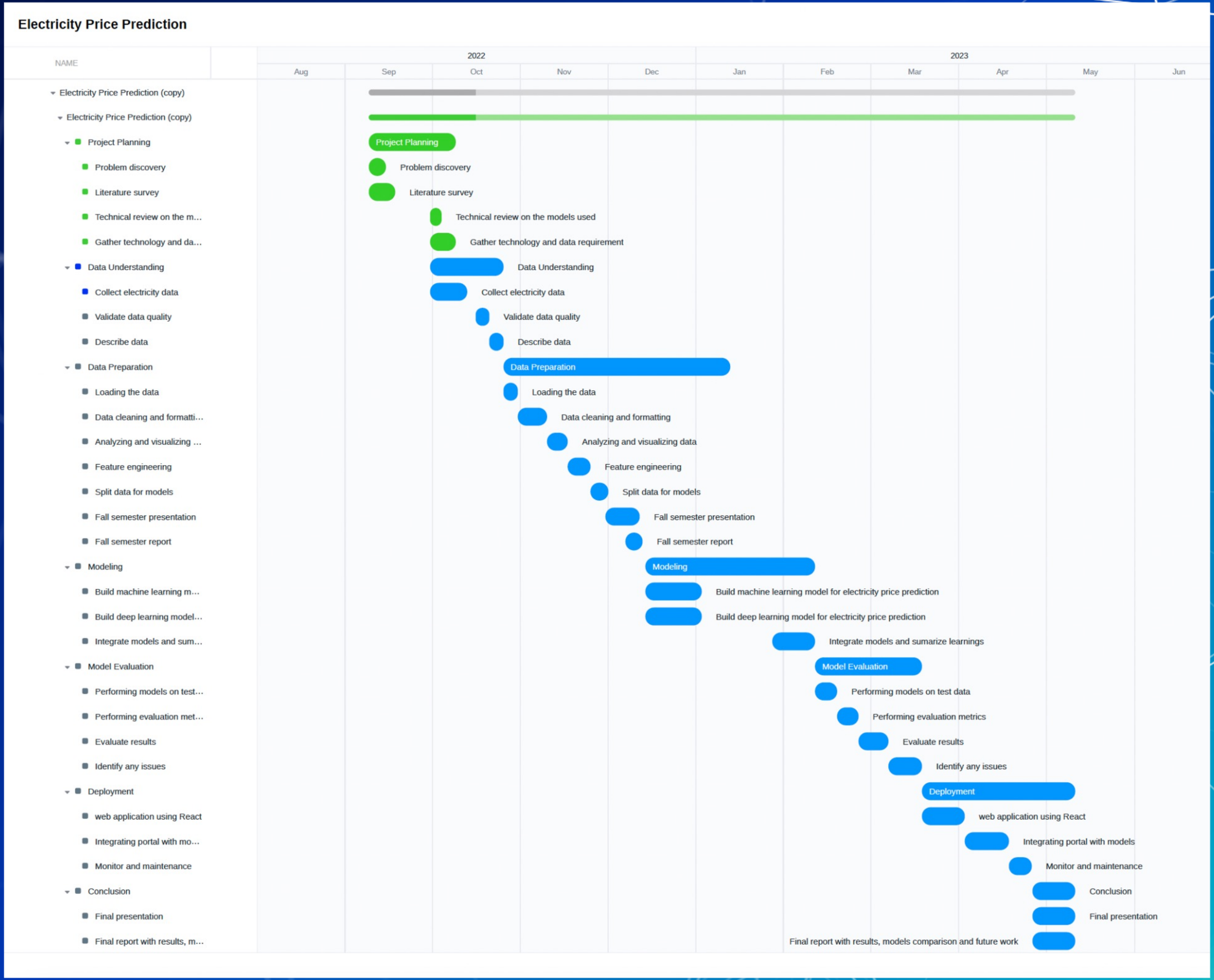


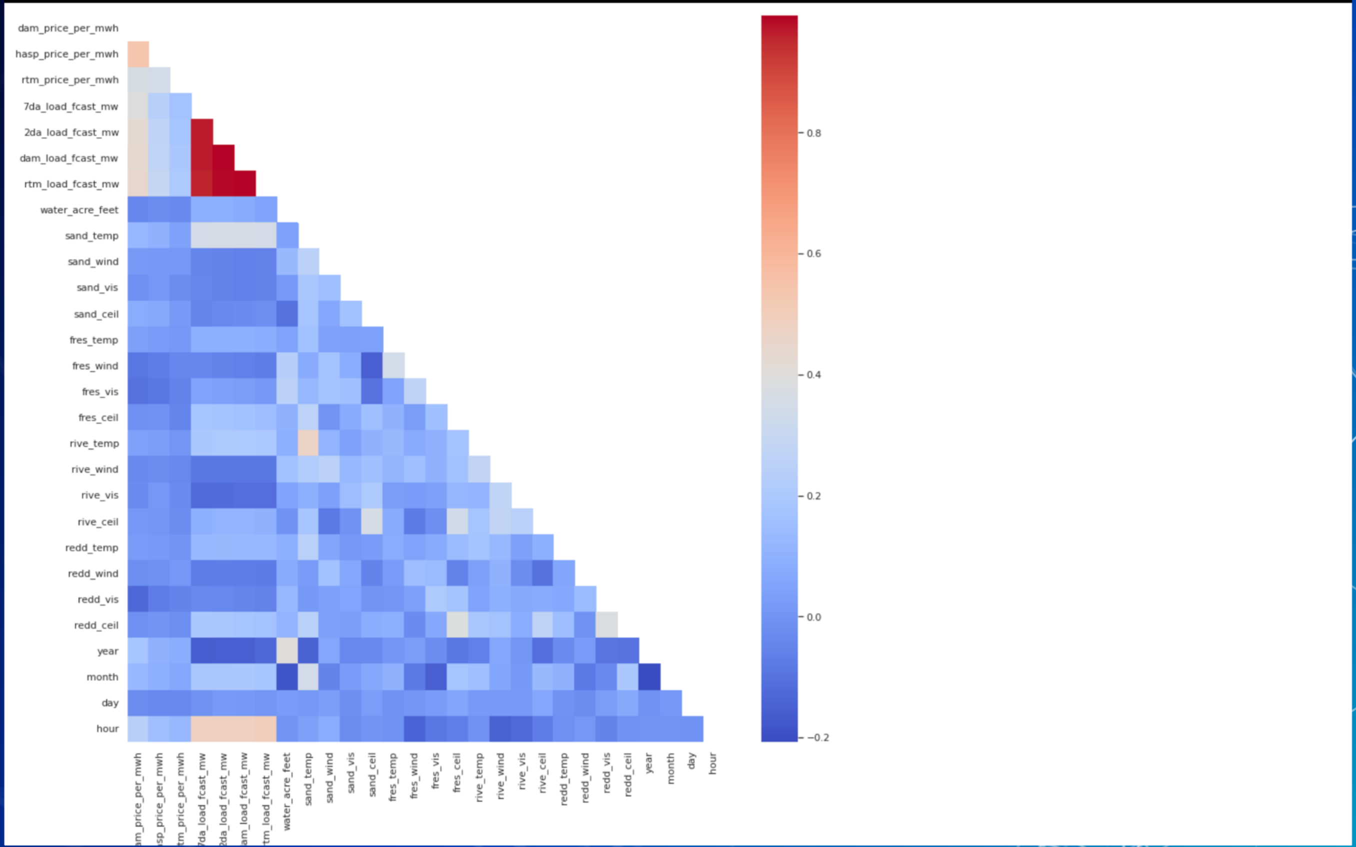



```
demand_data.describe()
```

	LOAD_TYPE	OPR_HR	OPR_INTERVAL	POS	MW	GROUP
count	1.113779e+07	1.113779e+07	1.113779e+07	1.113779e+07	1.113779e+07	1.113779e+07
mean	1.450646e-01	1.249623e+01	4.422373e+00	4.771313e+00	3.984654e+03	2.129754e+03
std	4.697968e-01	6.922711e+00	3.821788e+00	1.453188e+00	5.522388e+03	1.334884e+03
min	0.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	-8.029000e+03	1.000000e+00
25%	0.000000e+00	6.000000e+00	1.000000e+00	4.800000e+00	9.654300e+02	1.025000e+03
50%	0.000000e+00	1.200000e+01	4.000000e+00	5.300000e+00	2.251240e+03	1.984000e+03
75%	0.000000e+00	1.800000e+01	8.000000e+00	5.800000e+00	4.580220e+03	3.042000e+03
max	2.000000e+00	2.500000e+01	1.200000e+01	5.800000e+00	5.191410e+04	5.487000e+03

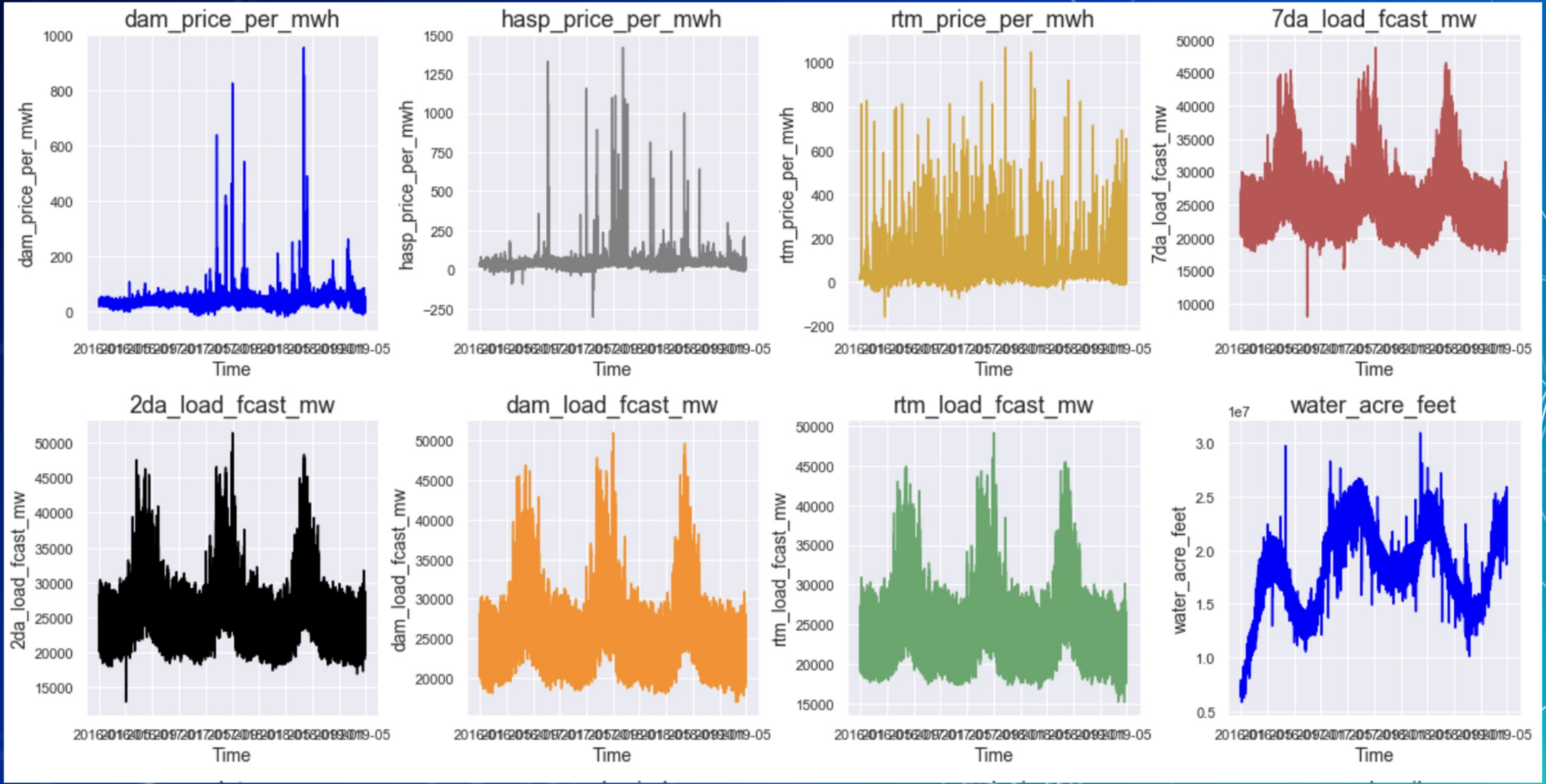
Gantt Chart


Electricity Price Prediction
NAME
2022
2023
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Electricity Price Prediction (copy)
Electricity Price Prediction (copy)
Project Planning
Problem discovery
Literature survey
Technical review on the m…
Gather technology and da…
Data Understanding
Collect electricity data
Validate data quality
Describe data
Data Preparation
Loading the data
Data cleaning and formatti…
Analyzing and visualizing …
Feature engineering
Split data for models
Fall semester presentation
Fall semester report
Modeling
Build machine learning m…
Build deep learning model…
Integrate models and sum…
Model Evaluation
Performing models on test…
Performing evaluation met…
Evaluate results
Identify any issues
Deployment
web application using React
Integrating portal with mo…
Monitor and maintenance
Conclusion
Final presentation
Final report with results, m…
Project Planning
Problem discovery
Literature survey
Technical review on the models used
Gather technology and data requirement
Data Understanding
Collect electricity data
Validate data quality
Describe data
Data Preparation
Loading the data
Data cleaning and formatting
Analyzing and visualizing data
Feature engineering
Split data for models
Fall semester presentation
Fall semester report
Modeling
Build machine learning model for electricity price prediction
Build deep learning model for electricity price prediction
Integrate models and sumarize learnings
Model Evaluation
Performing models on test data
Performing evaluation metrics
Evaluate results
Identify any issues
Deployment
web application using React
Integrating portal with models
Monitor and maintenance
Conclusion
Final presentation
Final report with results, models comparison and future work



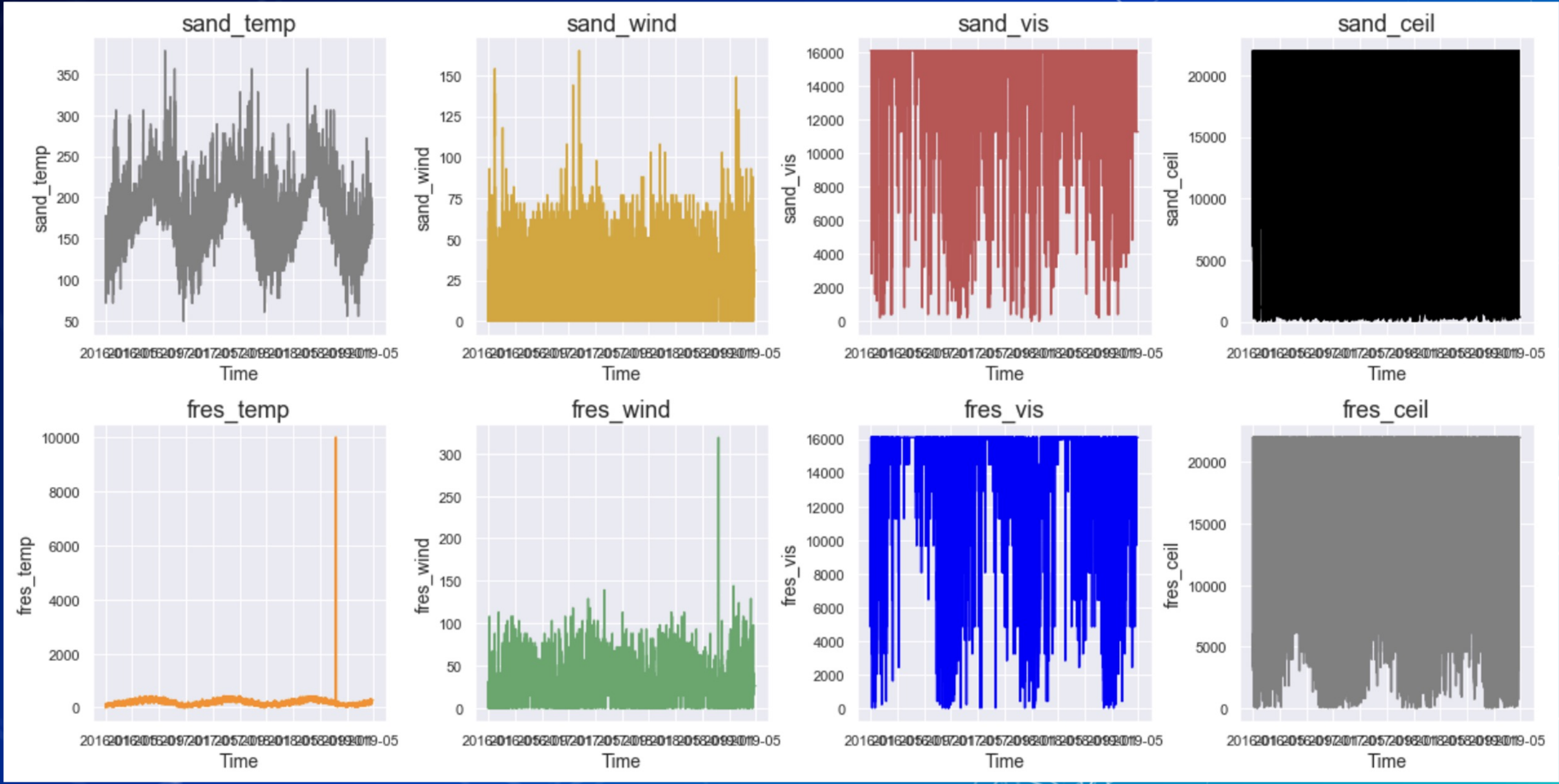
- More than 30,000 hours of data captured
- Overall correlations are weak

Data Visualization of Training data using gen_linecharts



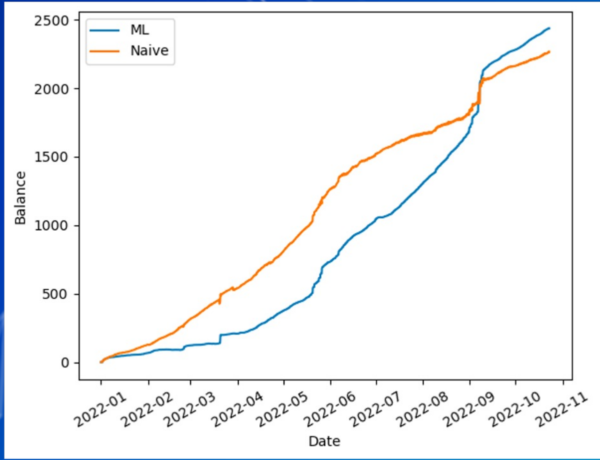
Two target columns(DAM and HASP)

Other features such as water , sand, wind etc



Simulation of a Tesla PowerWall

- Capacity: 13.5 kWh
- Charging/Discharge speed: 3.3 kW
- Assume we can sell at PGE's rate

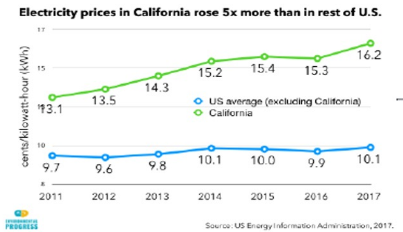


HOUR	OPR_HR	POS	VALUE
0	9.682261	1.0	-0.263018
1	10.682261	1.0	-0.308321
2	11.682261	1.0	-0.303699
3	12.682261	1.0	-0.293352
4	13.682261	1.0	-0.264003
5	14.682261	1.0	-0.202711
6	15.682158	1.0	-0.159727
7	16.681951	1.0	-0.068144
8	17.681951	1.0	0.126429
9	18.681951	1.0	0.414240
10	19.681951	1.0	0.443605
11	20.681951	1.0	0.336115
12	21.681951	1.0	0.216757
13	22.681951	1.0	0.164745
14	23.681951	1.0	0.094402
15	8.315122	1.0	0.054583
16	1.681287	1.0	0.017020
17	2.724172	1.0	-0.003407
18	3.682261	1.0	-0.012915
19	4.682261	1.0	0.001882
20	5.682261	1.0	0.040419
21	6.682261	1.0	0.067370
22	7.682261	1.0	0.013327
23	8.682261	1.0	-0.109909

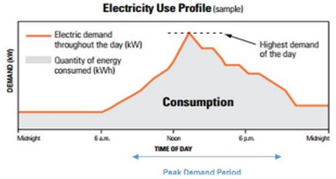
Machine Learning Results

Goal of the Model

©HR_SJSU



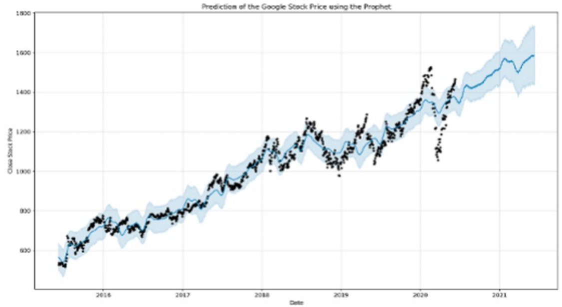
Past Prices



Past Demand

Other Features

Machine Learning Model



Future Prices

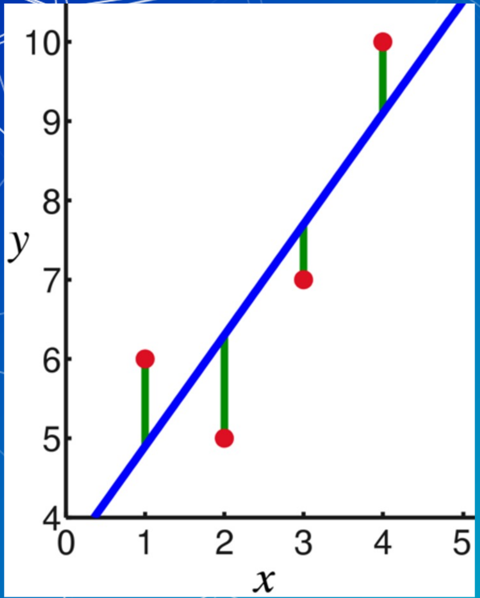
©HR_SJSU

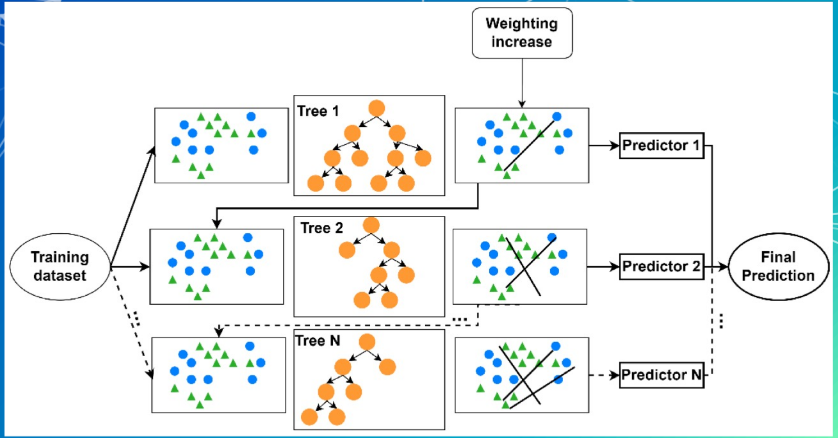
Model Inputs and Outputs

- Input
  - Past Data (hourly in the past 7 days)
    - Market price
    - Weather (temperature, moisture, etc)
    - Location, Event, Date, Time
  - Forecast Data (hourly in the next 7 days)
    - Demand and supply forecast
- Output
  - Predicted price in the next day (hourly interval)

Classic Machine Learning Models

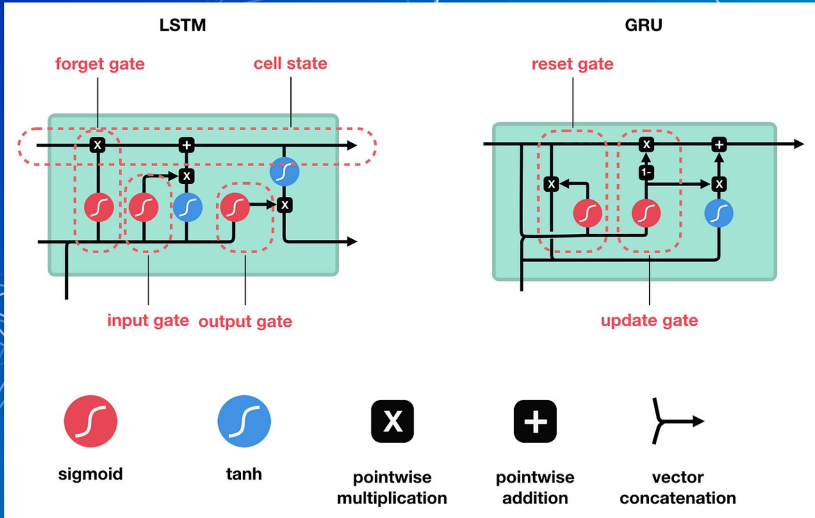
- We selected some classic machine learning algorithms as baselines
  - Linear Regression
  - XGBoost





Deep Learning Models

- Recurrent Neural Networks
  - Long-Short Term Memory
  - Gated Recurrent Unit



Innovative Machine Learning Models

- Feature Interaction
    - We leveraged embedding to model events and temporal features
    - Motivated by DCN, we add feature interaction modules
- Multi-Task Learning
    - We improved the model via adding multiple objectives instead of one price prediction
- Transformer
    - Leveraged Transformer Encoders

Feature Interaction

- Feature Interactions Between Prices and Weather/Events
- No manual feature engineering

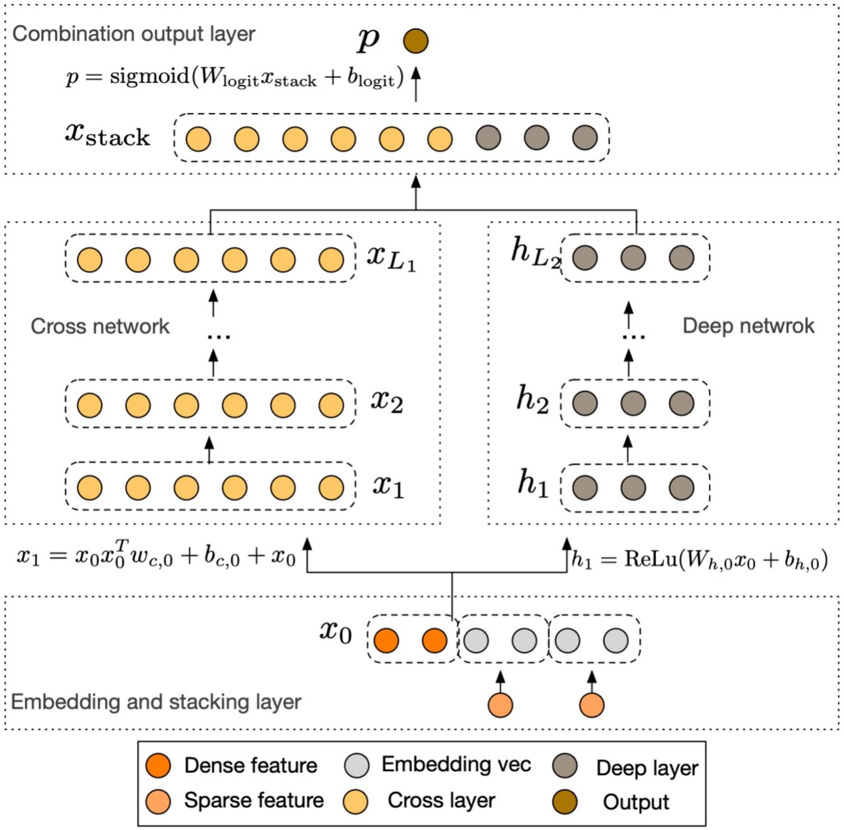


Figure 1: The Deep & Cross Network

Multi-Task Learning

- Market Price Prediction
- High and low prediction
- ...
- MTL improves overall learning

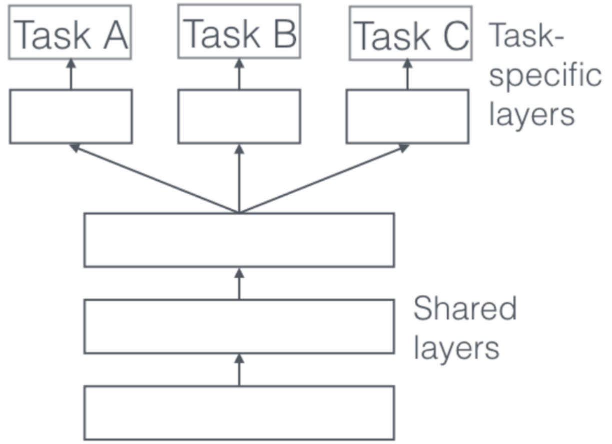
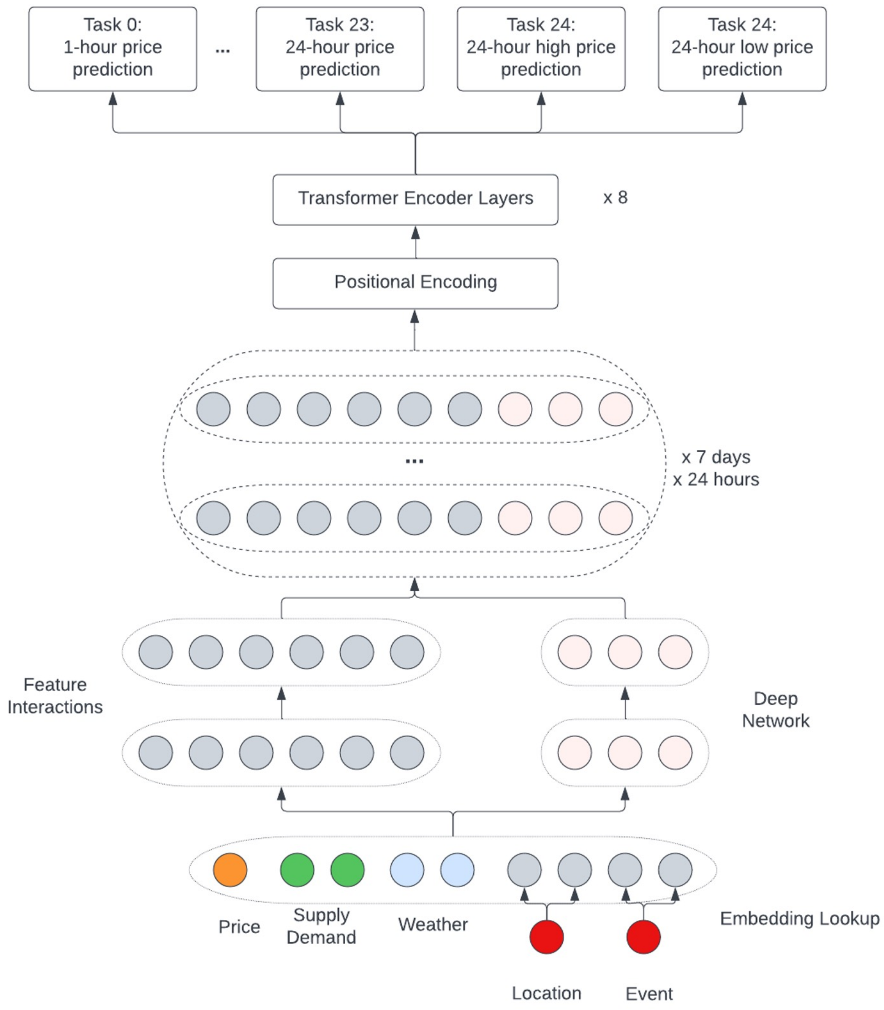
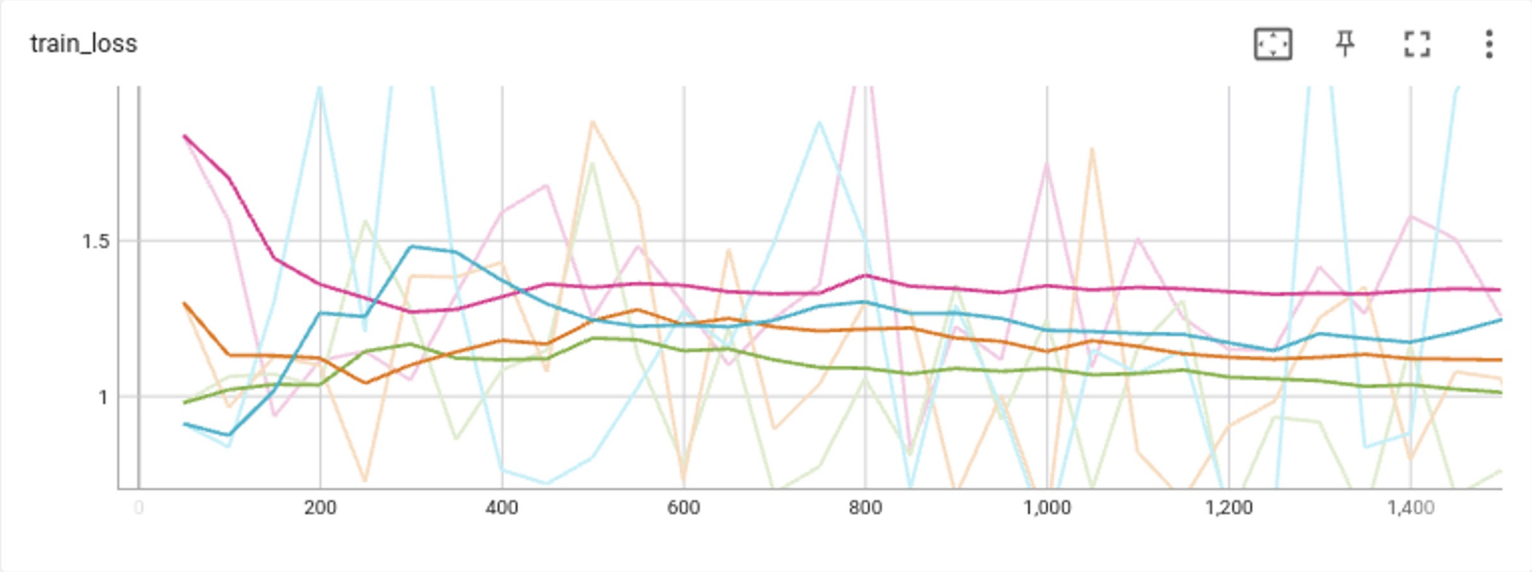


Figure 1: Hard parameter sharing for multi-task learning in deep neural networks

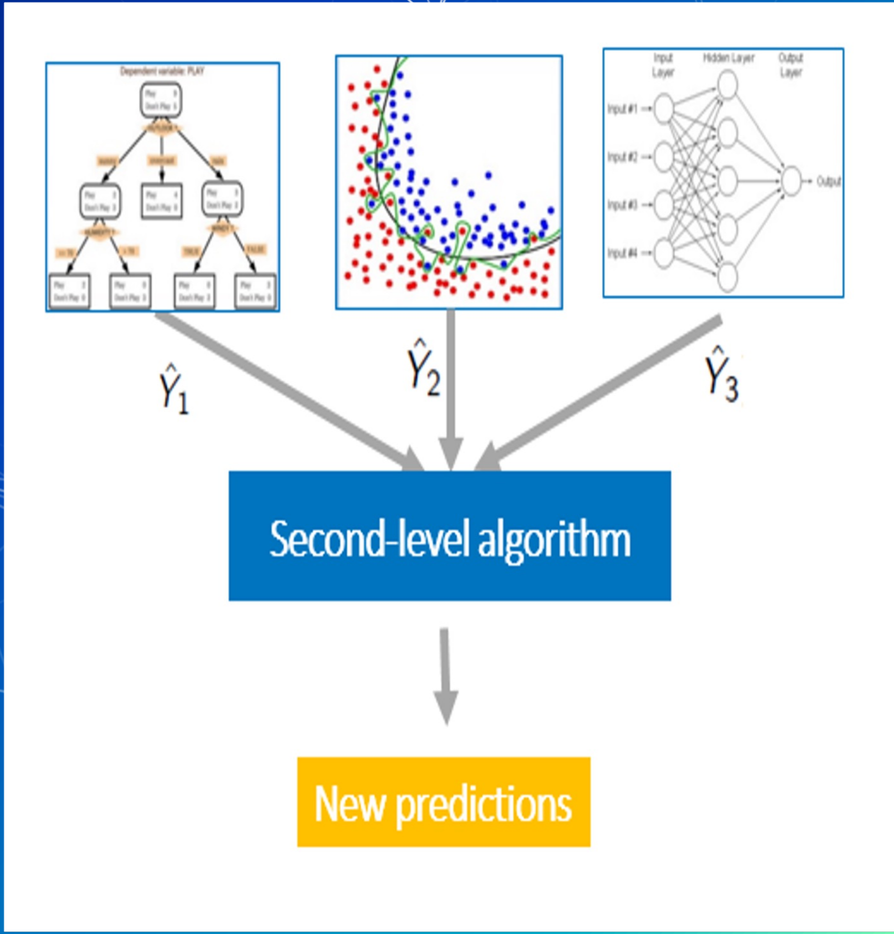

Task 0:
1-hour price
prediction
...
Task 23:
24-hour price
prediction
Task 24:
24-hour high price
prediction
Task 24:
24-hour low price
prediction
Transformer Encoder Layers
x 8
Positional Encoding
...
x 7 days
x 24 hours
Feature
Interactions
Deep
Network
Price
Supply
Demand
Weather
Embedding Lookup
Location
Event

Training Losses

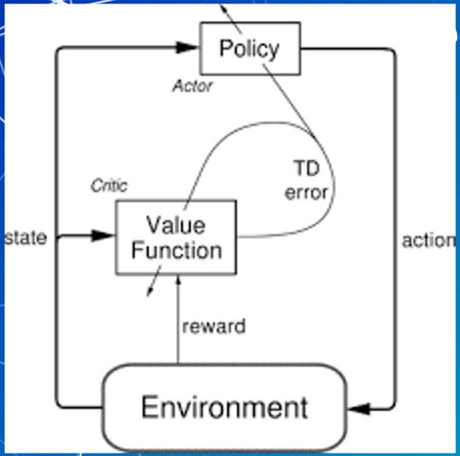


Ensemble Models

- Averaging everything together
  - It doesn't work!
  - Average is still worse than the best model.
- Learned weights based on inputs
  - Learned weights to combine model outputs
- StackingRegressor with RidgeRegression



Reinforcement Learning



- OpenAI Gym + Stable Baseline 3
- Action Space
  - Price ranges
- Observation Space
  - Features
- Reward
  - MSE loss
- Proximal Policy Optimization (PPO)
- MlpPolicy (ActorCriticPolicy)

Evaluation Metrics

- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)

Seasonality Evaluation

- All days of the week
- Weekdays (Mon - Fri)
- Summer Months (May - sep)

RMSE

Model	All days	Weekdays	Summer Months
Linear Regression	1.065	1.123	1.157
XGBoost	1.048	1.104	1.113
Reinforcement Learning	1.133	1.211	1.330
Ensemble	1.065	1.123	0.849
GRU	0.135	0.197	0.367
LSTM	0.117	0.410	0.281
Transformer	0.103	0.129	0.162
Proposed	0.057	0.578	0.271

MAE

Model	All days	Weekdays	Summer Months
Linear Regression	0.479	1.123	0.436
XGBoost	0.495	1.104	0.437
Reinforcement Learning	0.513	0.565	0.633
Ensemble	0.48	0.493	0.409
GRU	0.124	0.197	0.356
LSTM	0.117	0.405	0.272
Transformer	0.088	0.109	0.124
Proposed	0.049	0.075	0.118

MAPE

Model	All days	Weekdays	Summer Months
Linear Regression	1.562	1.123	1.634
XGBoost	2.790	1.104	3.082
Reinforcement Learning	4.213	4.386	4.707
Ensemble	1.408	1.012	1.624
GRU	0.555	0.197	2.792
LSTM	0.420	17.051	1.211
Transformer	0.275	0.462	0.396
Proposed	0.149	0.791	2.008

Runtime Comparison

Google Colab VM

GPU Model: Nvidia K80 / T4, GPU Memory: 12GB

GPU Memory Clock: 0.82GHz / 1.59GHz

Performance: 4.1 TFLOPS / 8.1 TFLOPS

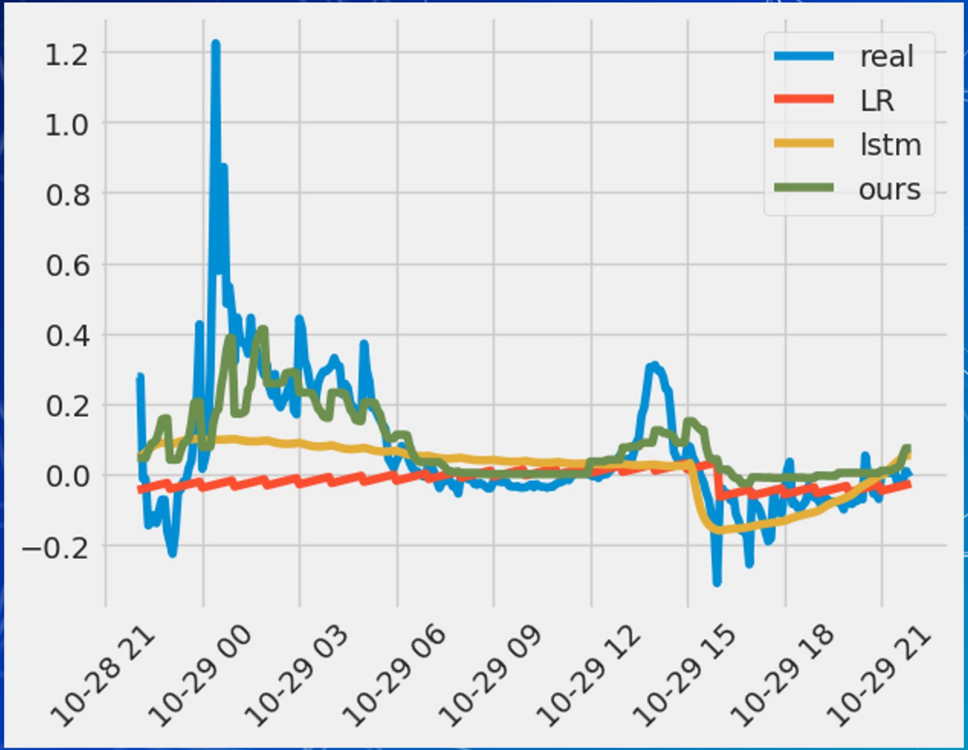
No. of CPU Cores: 2, RAM: 12GB

Model	Runtime in second
Linear Regression	0.058
XGBoost	10.54
GRU	1216
LSTM	219
Transformer	0.14
Proposed	1807.21
Proposed Optimized	301.3

Proposed Model
Training Cost Analysis

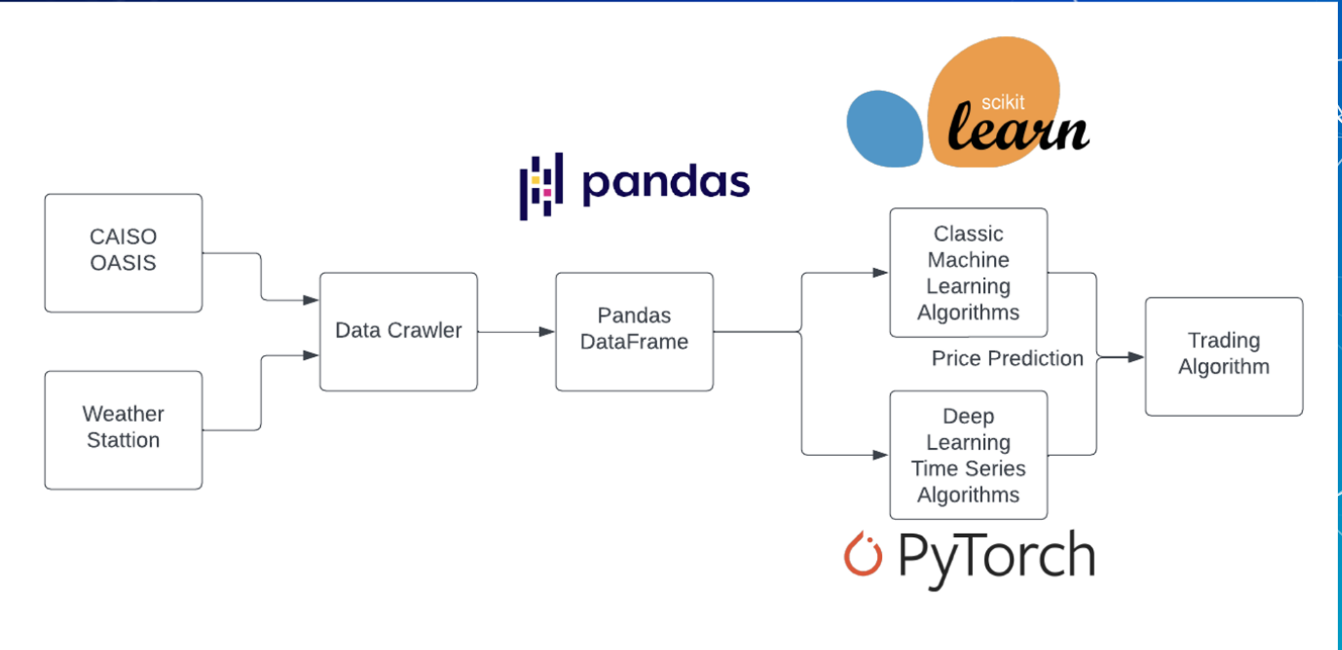
- Sparse lookups ~ 5%
- MLPs ~ 5%
- Transformer ~ 90%
  - Shows how transformer can leverage the power of computes
  - The optimization is mostly on this part to reduce the size of inputs

Actual Price vs Predicted Price

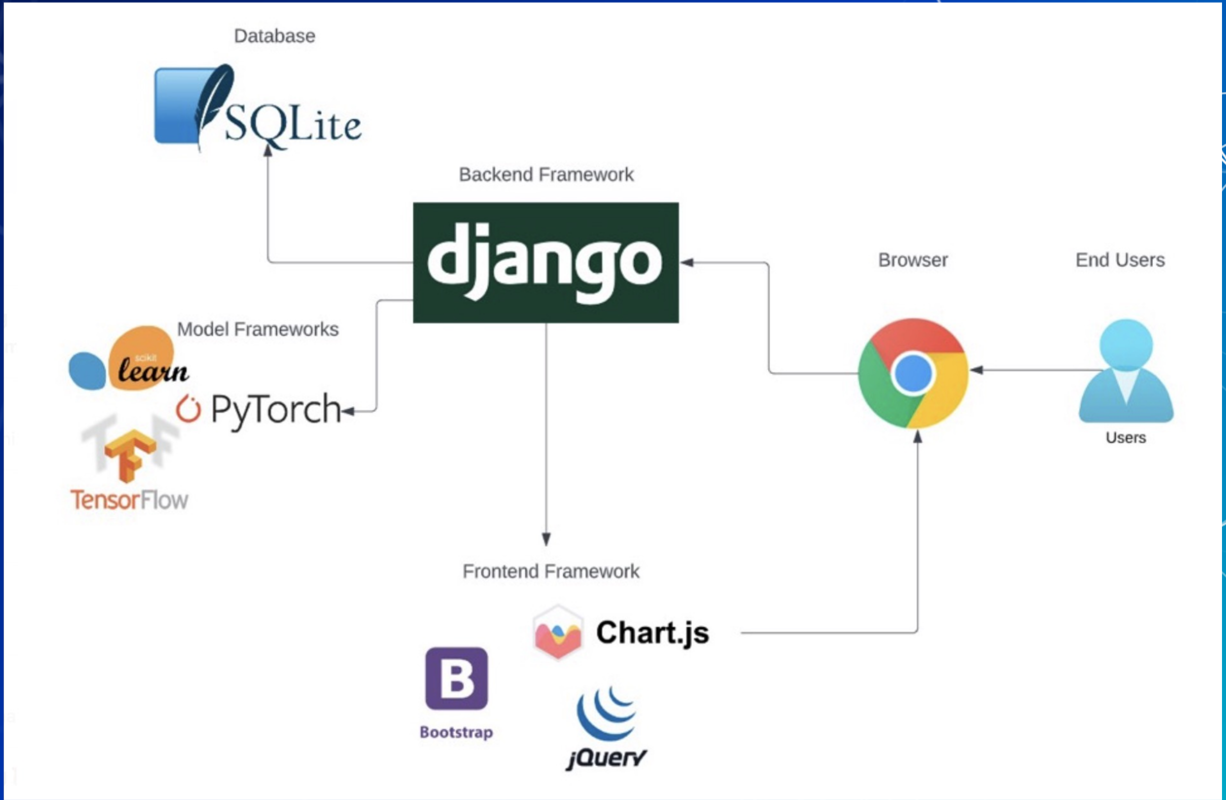


System Portal Design

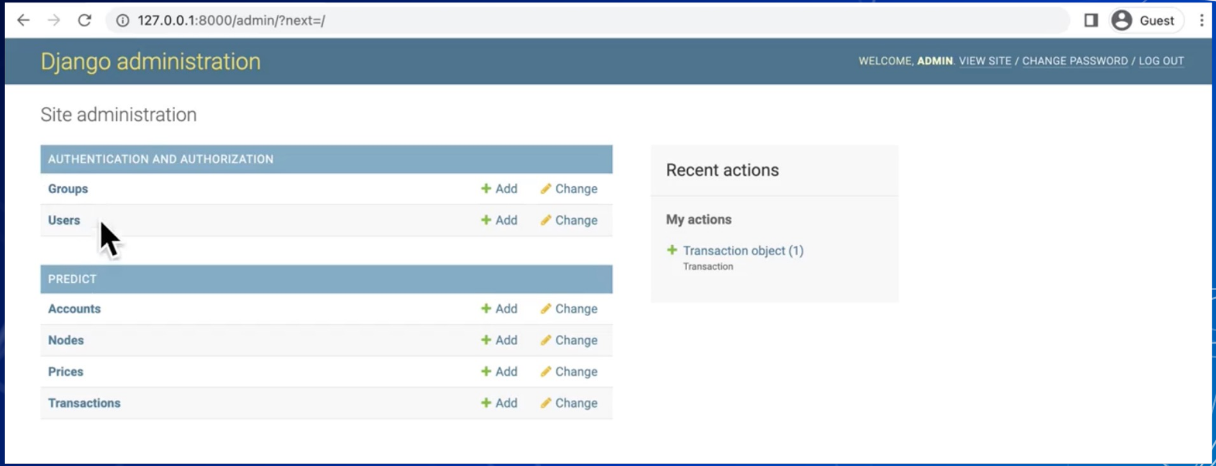
Model System Design



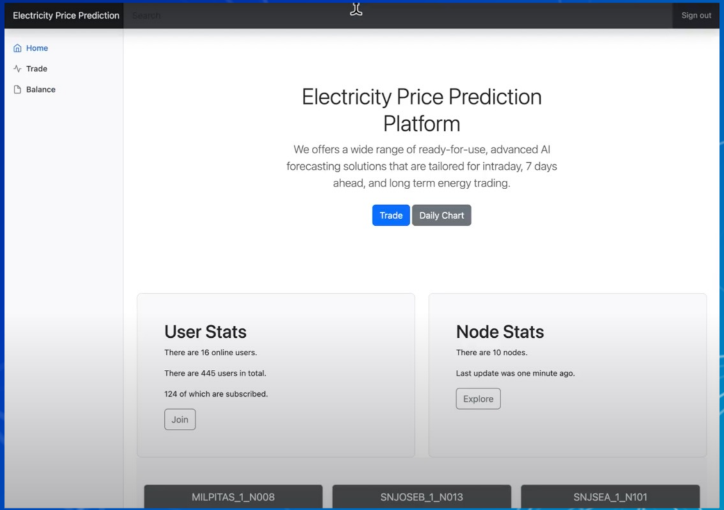
UI System Design

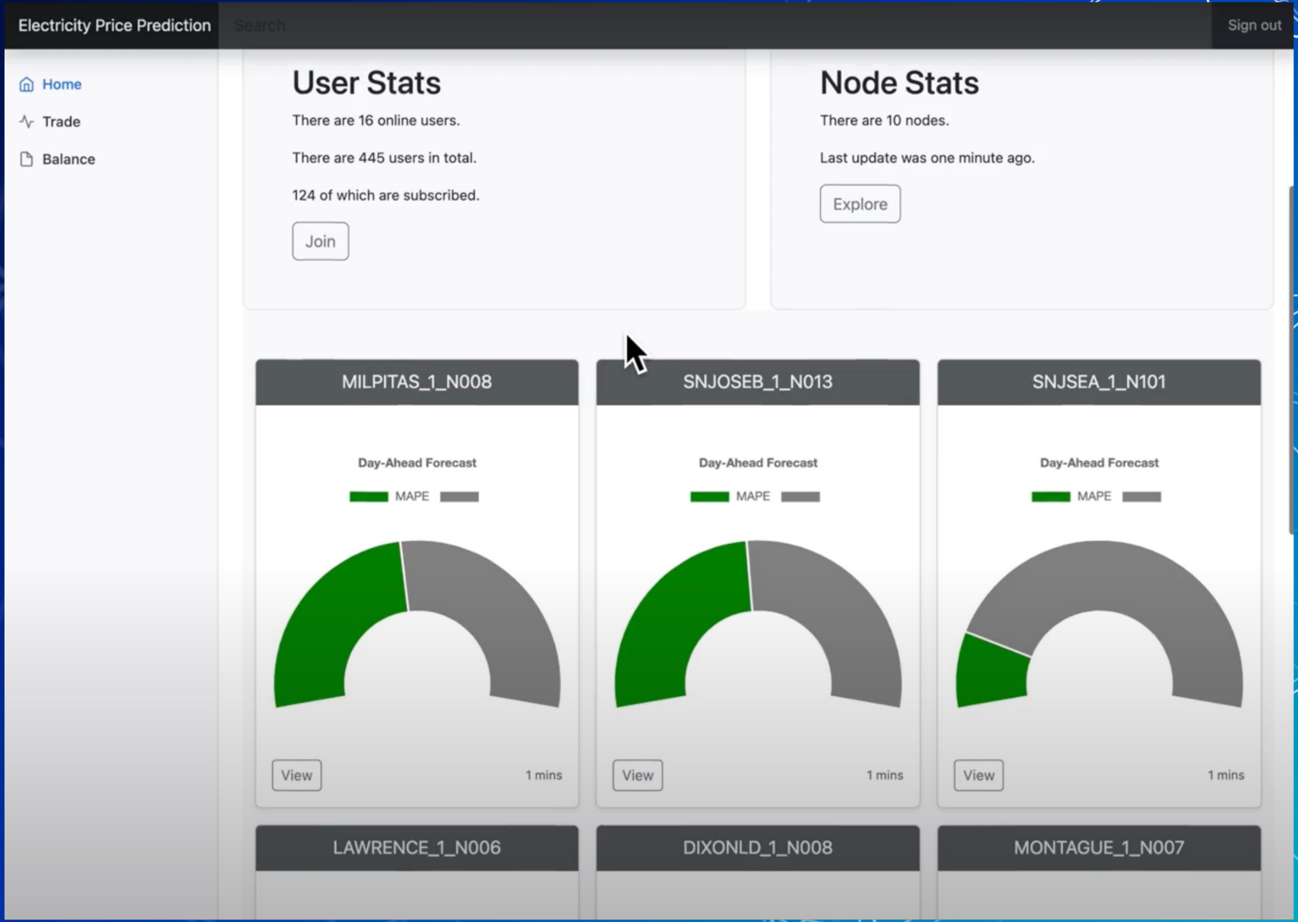


Admin Page

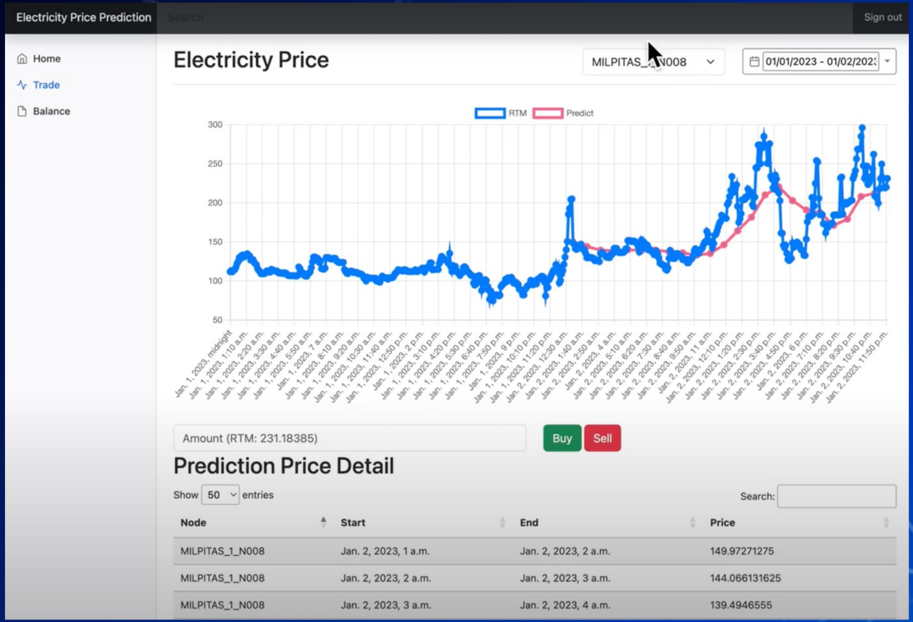


User Home page

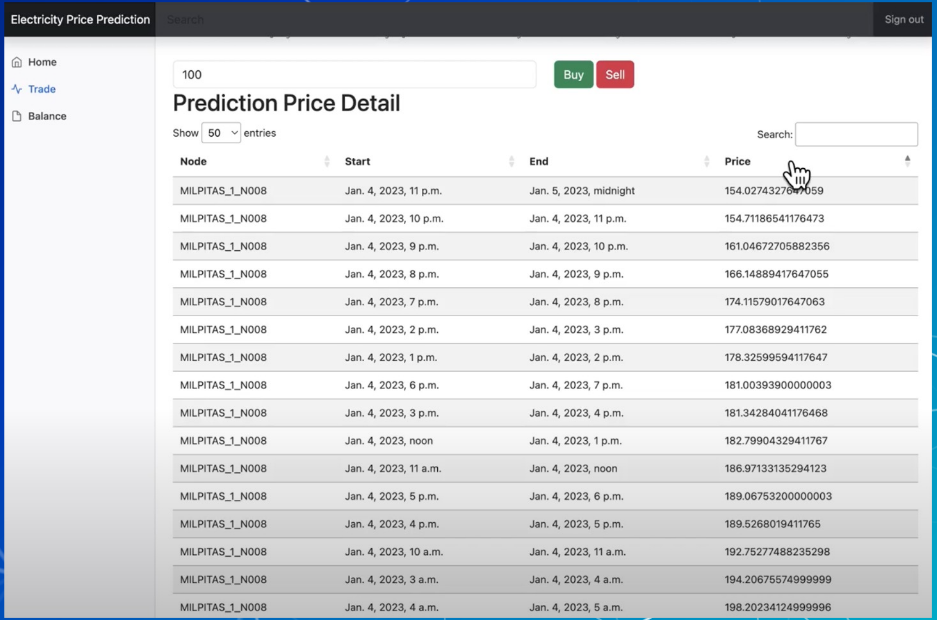




User interface with user and node statistics

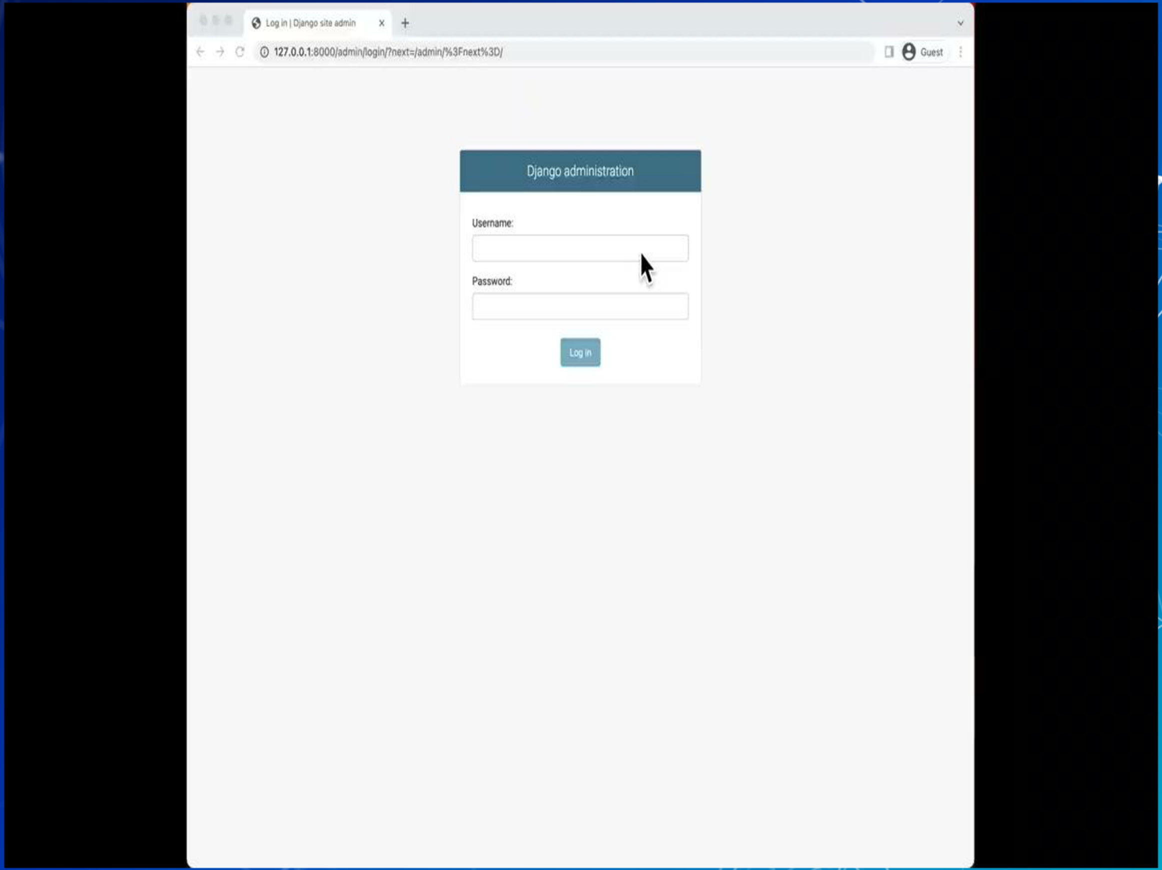


Demand and prediction graph



Prediction price detail on hourly basis

DEMO VIDEO


Log in | Django site admin
127.0.0.1:8000/admin/login/?next=/admin/%3Fnext%3D/
Guest
Django administration
Username:
Password:
Log in

THANK YOU