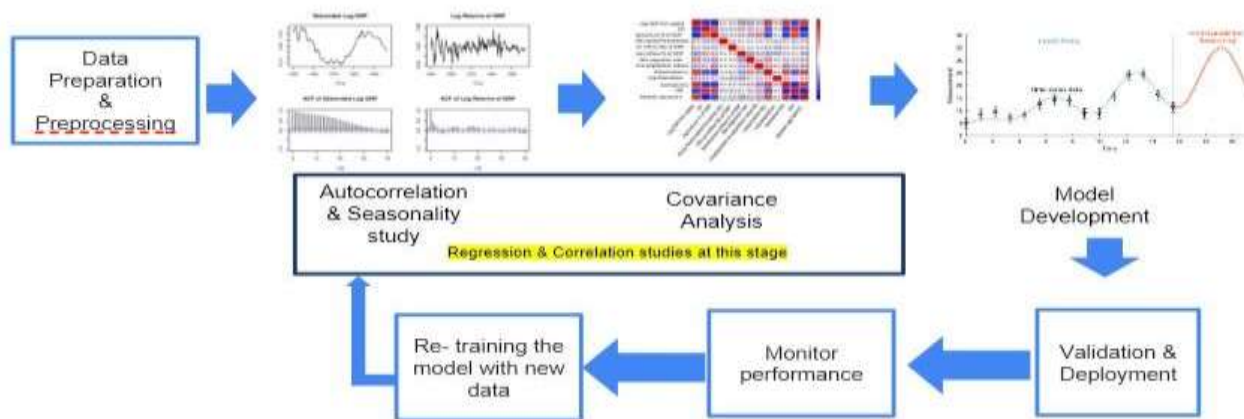


Problem Statement: *Social Impact Analysis using NTL data on various socio-economic factors.*

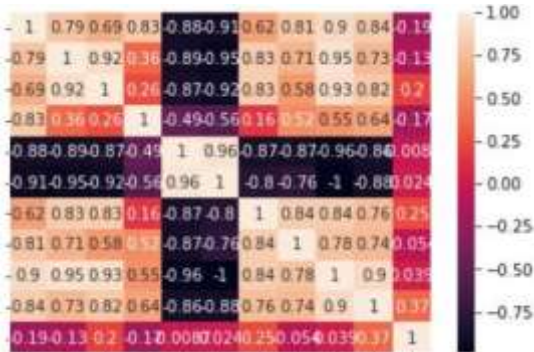
Explain your understanding on Problem Statement: In this project, we aim to use NTL data to find its correlation with various socio-economic factors to try to understand how it affects urbanization trends, road infrastructure, public health expenditure etc. This data also determines the economic status of rural areas.

Brief about your approach: To study the impacts of NTL on Socio economic parameters and understand the correlations and build forecasting models on the time series data. Combining the inputs from the NTL data and that from RBI, MOSPI and state govts, we attempt to build forecasting models. Model provides the insights and plan for changes of migration trends and its impact based on NTL data.



DETAILED APPROACH: STEP I. Shortlisting Strongly Correlated Socio-economic Factors

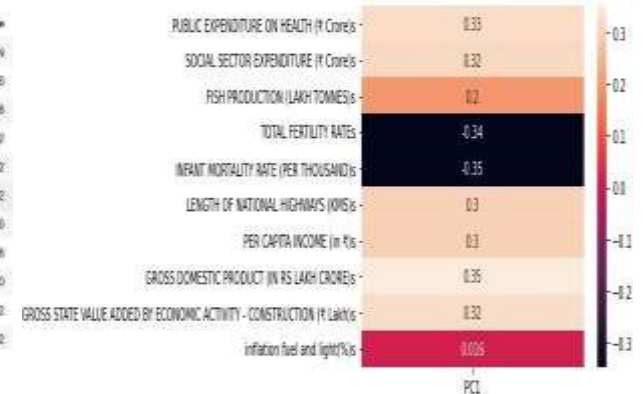
Correlation analysis



Covariance analysis

Parameter/Year	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020	covariance
NightTimeLight	31.43	30.93	31.10	32.34	33.48	34.28	NaN
PublicExpenditureHealth	2650.00	3948.00	5154.00	5213.20	5545.20	7303.80	5.128561e+06
SocialSectorExpenditure	24434.00	42510.00	47940.00	50204.00	56887.00	57910.00	5.182992e+08
FishProduction	2.68	2.37	1.99	2.70	2.84	3.00	4.105500e+02
TotalFertilityRate	1.80	1.80	1.70	1.70	1.60	1.60	4.315609e+02
InfantMortalityRate	34.00	31.00	29.00	27.00	23.00	21.00	8.905000e+02
NationalHighwayLength	2687.00	2696.00	3786.00	3786.00	3796.00	3796.00	3.538639e+06
PerCapitaIncome	124104.00	14080.00	159395.00	179358.00	209848.00	231378.00	8.715521e+08
GrossDomesticProduct	416332.00	464542.00	507946.00	557410.00	608401.00	640968.00	9.668811e+10
GSV_Construction	2333163.00	2398644.00	2435509.00	2784477.00	2905319.00	2759173.00	2.798139938916e+12
AverageInflationFuelLight	2.00	4.00	3.80	7.10	6.00	0.10	3.962749e+02

Principal Component analysis

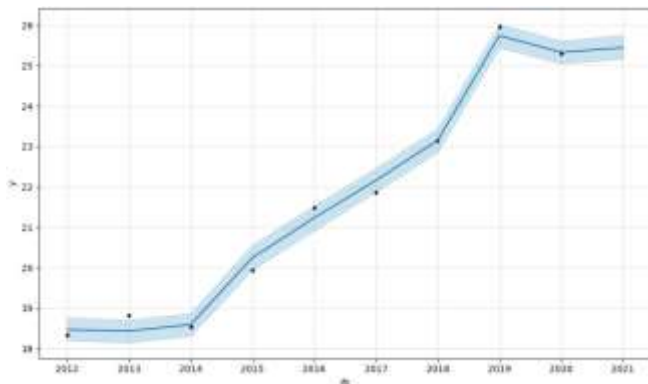


Shortlisted Socio-Economic Factors

	A	B	C	D	E	F	G	H	I	J
1	Parameter/Year	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020	Covariance	Correlation	PCA
2	NightTimeLight	31.43	30.93	31.1	32.34	33.48	34.28			
3	PerCapitaIncome	5128560.995	518299239.6	410.55	431.56095	0.08995	3538639.97	1.32906E+15	0.81	0.3
4	GSV_Construction	2333163	2398644	2435509	2784477	2905319	2759173	2798139938916	0.84	0.32
5	GrossDomesticProduct	416332	464542	507946	557410	608401	640968	96688111606	0.9	0.35
6	SocialSectorExpenditure	24434	42510	47949	50204	56887	57910	518299239.6	0.69	0.32
7	PublicExpenditureHealth	2650	3948	5154	5213.2	5545.2	7303.8	5128560.995	0.79	0.33
8	NationalHighwayLength	2687	2696	3786	3786	3796	3796	3538639.97	0.62	0.3
9	TotalFertilityRate	1.8	1.8	1.7	1.7	1.6	1.6	431.56095	-0.88	-0.34
10	FishProduction	2.68	2.37	1.99	2.7	2.84	3	410.55	0.83	0.2
11	AverageInflationFuelLight	2	4	3.8	7.1	6	0.1	396.27495	-0.19	0.016
12	InfantMortalityRate	34	31	29	27	23	21	0.08995	-0.91	-0.35

STEP II. Multivariate Time Series Regression Analysis Models - Top Correlated Socio-economic Factors

a.) FbProphet



Based on the previous years data from 2012-2020, multivariate regression is performed adding the following regressors: GDP, per capita income, Gross state value for road construction.

b.) DeepAR

```

) forecasts
[gluonts.model.forecast.SampleForecast(info=None, item_id=None, samples=array([[33.297844, 33.06234 ]], dtype=float32), start_date=Period('2018', 'A-DEC')),
gluonts.model.forecast.SampleForecast(info=None, item_id=None, samples=array([[207017.08, 171749. ]], dtype=float32), start_date=Period('2018', 'A-DEC')),
gluonts.model.forecast.SampleForecast(info=None, item_id=None, samples=array([[2832602., 4094406.]], dtype=float32), start_date=Period('2018', 'A-DEC')),
gluonts.model.forecast.SampleForecast(info=None, item_id=None, samples=array([[781283.2 , 558653.06]], dtype=float32), start_date=Period('2018', 'A-DEC')),
gluonts.model.forecast.SampleForecast(info=None, item_id=None, samples=array([[55741.137, 59634.19 ]], dtype=float32), start_date=Period('2018', 'A-DEC')),
gluonts.model.forecast.SampleForecast(info=None, item_id=None, samples=array([[5817.221, 5866.344]], dtype=float32), start_date=Period('2018', 'A-DEC'))]
```

According to DeepAR model, the predictions for the future years is made through multivariate regression method.

Expected outcomes

Establish forecasting models for Socio-Economic parameters based on their correlations & also NTL predications to cross validate.

$$\text{GDP}(t) = a * \text{NTL}(t-1) + b * (\text{Farming and/or employment and /or electricity consumption}) + \text{Tracking Error}$$

$$\text{NTL}(t) = a1 * \text{GDP}(t-1) + b1 * (\text{Farming and / or Roads and / or employment and /or electricity consumption}) + \text{Tracking Error}$$

Tools/models/APIs and devices used on development

- **Google Colaboratory**
- **FbProphet**
- **DeepAR**
- **Vector AR**
- **Python**

Technologies involved/used

- Machine learning/Deep learning
- Data science

References/Acknowledgement

- *Han G, Zhou T, Sun Y, Zhu S.* The relationship between night-time light and socioeconomic factors in China and India. PLoS One. 2022 Jan 13;17(1):e0262503. doi: 10.1371/journal.pone.0262503. PMID: 35025972; PMCID: PMC8758086.
- *Lu Y, Coops NC (2018)* Bright lights, big city: Causal effects of population and GDP on urban brightness. PLoS ONE 13(7): e0199545. doi: 10.1371/journal.pone.0199545