

AI-Driven Exploration and prediction of company registration trends with registrar of companies(ROC)

510521205033:Renuka H

Bharathidasan Engineering College

Project Title:ROC company analysis



Introduction:

In an era characterized by unprecedented growth in business and evolving regulatory landscapes, the Registrar of Companies (RoC) stands as a pivotal institution responsible for overseeing and regulating corporate entities. The RoC plays a crucial role in maintaining transparency, ensuring compliance, and fostering an environment conducive to economic development. As the volume of business registrations continues to surge, the need for

sophisticated analytical tools and predictive insights has never been more pressing.

The Registrar of Companies (ROC) plays a pivotal role in the governance and regulation of corporate entities within a country. Its mandate involves the registration, management, and oversight of businesses, ensuring compliance with statutory requirements and facilitating transparency in the corporate landscape. In this era of data-driven decision-making, conducting a comprehensive ROC Company Analysis emerges as a valuable and insightful exercise.

ROC Company Analysis involves the systematic examination of data and information provided by the Registrar of Companies pertaining to registered businesses. This data encompasses a wide array of details, including but not limited to, company registrations, financial statements, ownership structures, and compliance records. Through the prism of analytics and data science, this analysis endeavors to extract meaningful insights and predictions regarding company behaviors, industry trends, and regulatory compliance.

This project represents a fusion of technology, data science, and regulatory insight, offering a valuable resource for all stakeholders interested in the past, present, and future of company registrations within the purview of the Registrar of Companies.

Data Source:

The data source for ROC (Registrar of Companies) company analysis typically includes datasets and information provided by the Registrar of Companies or government authorities responsible for corporate registrations.

datasetLink:<https://tn.data.gov.in/resource/company-master-data-tamil-nadu-upto-28th-february-2019>

401	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con08-09-1988	Taril Nadi	50000000	30750000	1701	Manufacturer34 AKAMARU R O C	SECRETARY@3-03-2009	3-03-2009
402	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 5/4/1984	Taril Nadi	80000000	80000000	1701	Manufacturer No 1 Sunda R O C	Investment@3-03-2008	3-03-2008
403	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 06/10/1989	Taril Nadi	97500000	67500000	1701	Manufacturer5700 MARU R O C	Indonesian@3-03-2009	3-03-2009
404	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 10/7/1989	Taril Nadi	165000000	57542690	1701	Manufacturer252 METU R O C	Indonesian@3-03-2009	3-03-2009
405	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 06/10/1989	Taril Nadi	500000000	453233350	1701	Manufacturer6100 MARU R O C	Companysec@3-03-2004	3-03-2004
406	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 9/2/1990	Taril Nadi	200000000	194522000	1701	Manufacturer4100 MARU R O C	Indonesian@3-03-2002	3-03-2002
407	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 10/3/1993	Taril Nadi	108000000	60000000	1701	Manufacturer2310 MARU R O C	Indonesian@3-03-2009	3-03-2009
408	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 08-08-1994	Taril Nadi	600000000	52624000	1701	Manufacturer9F NO 74/12 R O C	Indonesian@3-03-2009	3-03-2009
409	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 07-03-1994	Taril Nadi	100000000	47888000	1701	Manufacturer3A B BLOCK R O C	Indonesian@3-03-2009	3-03-2009
410	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 09-03-2003	Taril Nadi	550000000	302803920	1701	Manufacturer9 GOKIL BUKU R O C	Indonesian@3-03-2009	3-03-2009
411	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 24-8-2005	Taril Nadi	150000000	86480000	1701	Manufacturer252 METU R O C	Indonesian@3-03-2009	3-03-2009
412	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 7/3/1980	Taril Nadi	193000000	192864000	1701	ManufacturerDor No 18 R O C	Indonesian@3-03-2009	3-03-2009
413	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 6/10/1988	Taril Nadi	100000000	57250000	1701	Manufacturer NO 9 ANUL R O C	Indonesian@3-03-2009	3-03-2009
414	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 26/07/1989	Taril Nadi	1225+09	10675+09	1701	Manufacturer25 25 A VEN R O C	Indonesian@3-03-2009	3-03-2009
415	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 25-06-1992	Taril Nadi	400000000	256378250	1701	Manufacturer THEN THRU R O C	Indonesian@3-03-2009	3-03-2009
416	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 22-08-1957	Taril Nadi	200000000	12899000	1701	Manufacturer NSCSE TONG R O C	Indonesian@3-03-2009	3-03-2009
417	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 05-8-1994	Taril Nadi	70000000	60439500	1701	ManufacturerLAND MARU R O C	Indonesian@3-03-2008	3-03-2008
418	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 10-12-2007	Taril Nadi	1443E+09	1430E+09	1701	ManufacturerNO 4 KARU R O C	Indonesian@3-03-2009	3-03-2009
419	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 4/4/1983	Taril Nadi	330000000	31260000	1701	ManufacturerNO 87 CILMER R O C	Indonesian@3-03-2003	3-03-2003
420	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 28-04-1988	Taril Nadi	800000000	728725050	1701	ManufacturerNOF - IV C R O C	Indonesian@3-03-2009	3-03-2009
421	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 07-07-1989	Taril Nadi	120000000	87900000	1701	ManufacturerWILLU R O C	Indonesian@3-03-2007	3-03-2007
422	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 02-02-1986	Taril Nadi	15000000	6674500	1701	Manufacturer3 GOMAR R O C	Indonesian@3-03-2009	3-03-2009
423	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 2/11/1988	Taril Nadi	1000000	2000	1701	Manufacturer5/1 MOU R O C	Indonesian@3-03-2009	3-03-2009
424	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 3/7/1990	Taril Nadi	300000000	25000000	1701	ManufacturerNo 4 Anul R O C	Indonesian@3-03-2009	3-03-2009
425	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 03-8-1994	Taril Nadi	500000000	15620000	1701	ManufacturerNEW NO 29 R O C	Indonesian@3-03-2009	3-03-2009
426	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 29-09-2009	Taril Nadi	120000000	75832600	1701	Manufacturer208 Vichet R O C	Indonesian@3-03-2009	3-03-2009
427	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 08-8-2005	Taril Nadi	472500000	256930000	1701	Manufacturer30 A EXTEN R O C	Indonesian@3-03-2009	3-03-2009
428	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 05-8-2001	Taril Nadi	400000000	26922700	1701	ManufacturerNUS MUR R O C	Indonesian@3-03-2009	3-03-2009
429	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 25-04-1998	Taril Nadi	500000	700	18019	ManufacturerNUS55 MARU R O C	Indonesian@3-03-2009	3-03-2009
430	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 5/2/1992	Taril Nadi	50000000	39080250	18019	ManufacturerNo 10 PP Anul R O C	Indonesian@3-03-2009	3-03-2009
431	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 02-12-1993	Taril Nadi	60000000	50087000	18019	ManufacturerPLOT N03 R O C	Indonesian@3-03-2009	3-03-2009
432	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 6/12/1991	Taril Nadi	64500000	6370000	18019	Manufacturer47 UO MARU R O C	Indonesian@3-03-2009	3-03-2009
433	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 09-1974	Taril Nadi	15000000	4000000	18019	Manufacturer45 STEEL R O C	Indonesian@3-03-2009	3-03-2009
434	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 24-09-1987	Taril Nadi	200000000	145000000	18019	Manufacturer87/84 MARU R O C	Indonesian@3-03-2009	3-03-2009
435	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 4/9/1990	Taril Nadi	650000000	335294800	20013	ManufacturerNo 572 Anul R O C	Indonesian@3-03-2009	3-03-2009
436	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 3/11/2005	Taril Nadi	500000000	43836500	20013	ManufacturerKIDANG R O C	Indonesian@3-03-2006	3-03-2006
437	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 22-06-1990	Taril Nadi	700000000	12636200	20013	ManufacturerPILIPALAY R O C	Indonesian@3-03-2009	3-03-2009
438	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 02-12-1991	Taril Nadi	100000000	95550000	20013	ManufacturerNO 25 50 R O C	Indonesian@3-03-2009	3-03-2009
439	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 22-08-1975	Taril Nadi	8000000	0	20013	ManufacturerTSNO2429 R O C	Indonesian@3-03-2009	3-03-2009
440	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 9/7/1991	Taril Nadi	60000000	0	20013	ManufacturerNS0707 R O C	Indonesian@3-03-2009	3-03-2009
441	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 24-08-1995	Taril Nadi	600000000	26296000	20013	Manufacturer28 22 ANUL R O C	Indonesian@3-03-2009	3-03-2009
442	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 02-12-1985	Taril Nadi	0E+09	197E+09	22019	ManufacturerMasak Maru R O C	Indonesian@3-03-2009	3-03-2009
443	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 11/3/1994	Taril Nadi	3000000	2000	22019	Manufacturer321 GIST R O C	Indonesian@3-03-2009	3-03-2009
444	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 04-04-1979	Taril Nadi	133E+09	69200000	22019	Manufacturer67 MOUNT R O C	Indonesian@3-03-2009	3-03-2009
445	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 08-08-1970	Taril Nadi	200000000	180890000	22019	Manufacturer62 TOWER R O C	Indonesian@3-03-2009	3-03-2009
446	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 09-09-1974	Taril Nadi	35000000	30800000	22019	Manufacturer752 ANUL R O C	Indonesian@3-03-2009	3-03-2009
447	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 22-06-1984	Taril Nadi	2E+09	89974740	23019	ManufacturerMANU EXPR R O C	Indonesian@3-03-2009	3-03-2009
448	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 1/7/1970	Taril Nadi	250000000	95554425	24019	ManufacturerKOTAWA R O C	Indonesian@3-03-2009	3-03-2009
449	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 23-12-1974	Taril Nadi	500000000	49022700	24019	ManufacturerFlat N01 R O C	Indonesian@3-03-2009	3-03-2009
450	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 09-09-1974	Taril Nadi	15000000	0	24019	ManufacturerINDIAN CHAIR R O C	Indonesian@3-03-2009	3-03-2009
451	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 20-12-1972	Taril Nadi	350000000	149750000	24019	Manufacturer44 SPOT N00 R O C	Indonesian@3-03-2009	3-03-2009
452	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 04-04-1970	Taril Nadi	250000000	134509004	24019	ManufacturerThe Bara G R O C	Indonesian@3-03-2005	3-03-2005
453	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 09-09-1974	Taril Nadi	35000000	17530000	24019	ManufacturerNO 2 SHREE R O C	Indonesian@3-03-2006	3-03-2006
454	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 05-05-1990	Taril Nadi	50000000	46743000	24019	Manufacturer7 A SEDU R O C	Indonesian@3-03-2004	3-03-2004
455	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 02-02-1995	Taril Nadi	100000000	700	24019	Manufacturer91 EN PRA R O C	Indonesian@3-03-2009	3-03-2009
456	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 3/12/1971	Taril Nadi	122E+09	128E+09	24019	ManufacturerSFC HSE R O C	Indonesian@3-03-2009	3-03-2009
457	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 09-09-1974	Taril Nadi	500000000	0	24019	Manufacturer11/201 R O C	Indonesian@3-03-2009	3-03-2009
458	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 03-03-1992	Taril Nadi	100000000	46789300	24019	ManufacturerUO N017 R O C	Indonesian@3-03-2008	3-03-2008
459	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 03-03-1992	Taril Nadi	25000000	1033000	24019	Manufacturer2381 DHU R O C	Indonesian@3-03-2009	3-03-2009
460	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 22-09-1975	Taril Nadi	437E+09	17099488	24019	Manufacturer TWEE HUSE R O C	Indonesian@3-03-2009	3-03-2009
461	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 23-04-1977	Taril Nadi	60000000	49360000	24222	ManufacturerKUBUK R O C	Indonesian@3-03-2009	3-03-2009
462	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 1/7/1992	Taril Nadi	15E+09	889643270	24222	Manufacturer OCHO TOMO R O C	Indonesian@3-03-2008	3-03-2008
463	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 7/2/1973	Taril Nadi	100000000	0	24230	Manufacturer10 WAPOR R O C	Indonesian@3-03-2009	3-03-2009
464	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 25-07-1980	Taril Nadi	200000000	13000000	24230	ManufacturerCHEMIE R O C	Indonesian@3-03-2008	3-03-2008
465	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 20-08-1985	Taril Nadi	99000000	700	24230	ManufacturerOLINO R O C	Indonesian@3-03-2009	3-03-2009
466	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 5/10/1934	Taril Nadi	350000000	0	24230	Manufacturer3 WEST BUKU R O C	Indonesian@3-03-2009	3-03-2009
467	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 9/9/1936	Taril Nadi	50000000	29230630	24230	Manufacturer N003 (LD R O C	Indonesian@3-03-2009	3-03-2009
468	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 02-05-1958	Taril Nadi	200000000	14303330	24230	ManufacturerNo 6 CATER R O C	Indonesian@3-03-2009	3-03-2009
469	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 1/11/1983	Taril Nadi	50000000	4531000	24230	ManufacturerPLOT N043 R O C	Indonesian@3-03-2009	3-03-2009
470	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 7/5/1984	Taril Nadi	745000000	108330000	24230	ManufacturerNO 439 CHAIR R O C	Indonesian@3-03-2009	3-03-2009
471	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 07-08-1985	Taril Nadi	75000000	49327700	24230	Manufacturer23DR AMBER R O C	Indonesian@3-03-2008	3-03-2008
472	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 9/4/1986	Taril Nadi	10000000	0	24230	Manufacturer70 LUNG R O C	Indonesian@3-03-2009	3-03-2009
473	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 9/2/1987	Taril Nadi	2400000	7000	24230	Manufacturer3 PLAT SAN R O C	Indonesian@3-03-2009	3-03-2009
474	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 4/3/1987	Taril Nadi	150000000	74520000	24230	ManufacturerREGENCY R O C	Indonesian@3-03-2009	3-03-2009
475	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 11/9/1987	Taril Nadi	85000000	84633500	24230	ManufacturerPLOT N0A R O C	Indonesian@3-03-2009	3-03-2009
476	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 07-06-1988	Taril Nadi	700000000	68290000	24230	Manufacturer3839 A 40 R O C	Indonesian@3-03-2007	3-03-2007
477	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 26-03-1990	Taril Nadi	150000000	0	24230	ManufacturerSATWARI R O C	Indonesian@3-03-2009	3-03-2009
478	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 04-04-1990	Taril Nadi	170000000	15287500	24230	ManufacturerNIRAWA R O C	Indonesian@3-03-2009	3-03-2009
479	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 04-02-1991	Taril Nadi	30000000	30000000	24230	ManufacturerHP MATH R O C	Indonesian@3-03-2009	3-03-2009
480	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 3/6/1991	Taril Nadi	150000000	0	24230	Manufacturer44 BECH R O C	Indonesian@3-03-2009	3-03-2009
481	U7012208080PDR SUBSACTV	PdMc	Company linNo-gort con 04-02-1992	Taril Nadi	155000000	13592000				

301	906296	SARCHEM INACTV	NA	NA	NA	26-12-2017	Tamil Nadu	0	0	0	NA	Agriculture & NO 75 ADDLR O C	ENNA	ravisar	NA	NA	
302	90630	DEEP DRILLING INEF	NA	NA	NA	16-02-2018	Tamil Nadu	0	0	0	NA	Agriculture & ANPRIYA R R O C	ENNA	munagappan	NA	NA	
303	90632	BRAND R INACTV	NA	NA	NA	5/3/2018	Tamil Nadu	0	0	0	NA	Agriculture & No 5 / 329 R O C	ENNA	anandume	NA	NA	
304	906349	TANVY CRUACTV	NA	NA	NA	18-06-2018	Tamil Nadu	0	0	0	NA	Agriculture & No No. GBC R O C	ENNA	kingen@	NA	NA	
305	906371	SOLUTION INACTV	NA	NA	NA	13-09-2018	Tamil Nadu	0	0	0	NA	Agriculture & DOOR NO83 R O C	ENNA	shree7@sk2	NA	NA	
306	906384	EASTERN GLUACTV	NA	NA	NA	22-10-2018	Tamil Nadu	0	0	0	NA	Agriculture & SF NO 262 / 28 O C	ENNA	avinachari	NA	NA	
307	90640	TRANSMILE ACTV	NA	NA	NA	24-12-2018	Tamil Nadu	0	0	0	NA	Agriculture & No 83 A W R O C	ENNA	chandru	NA	NA	
308	90648	Sunite TechACTV	NA	NA	NA	17-02-2019	Tamil Nadu	0	0	0	NA	Agriculture & The Royal GR O C	ENNA	senbil_kuma	NA	NA	
309	90644	HYDRONGHEACTV	NA	NA	NA	20-03-2019	Tamil Nadu	0	0	0	NA	Agriculture & Door No 78 R R O C	ENNA	bcjeong@	NA	NA	
310	906408	PODSUN OY ACTV	NA	NA	NA	17-06-2019	Tamil Nadu	0	0	0	NA	Agriculture & Site 36 DES O C	ENNA	kouskivaku	NA	NA	
311	906487	ThinkTech INACTV	NA	NA	NA	20-07-2019	Tamil Nadu	0	0	0	NA	Agriculture & Prince hoci R O C	ENNA	Vijay Venkai	NA	NA	
312	000772944	INEDAMALA INACTV	Public	Company	linNon-govt	con21-04-1943	Tamil Nadu	12500000	6273500			1117 Agriculture & KATARY EST R O C	ENNA	secretaria	3-03-2019	3-03-2019	
313	00091988	WAN OFFS INACTV	Public	Company	linNon-govt	con25-09-1986	Tamil Nadu	15E+10	10730000			1119 Agriculture & ANPRIYA R R O C	ENNA	secretaria	3-03-2019	3-03-2019	
314	00091988	ASUMI R R O C	Public	Company	linNon-govt	con29-03-1989	Tamil Nadu	150000000	38350000			1119 Agriculture & 29 POLICE CR O C	ENNA	NA	NA	NA	
315	00091988	2507TECH INACTV	Public	Company	linNon-govt	con21-12-1992	Tamil Nadu	150000000	40000000			1119 Agriculture & 29 DREGS R O C	ENNA	compliance	3-03-2018	3-03-2018	
316	00091988	FLORENCE MUJO	Public	Company	linNon-govt	con30-11-1993	Tamil Nadu	500000	0			1119 Agriculture & 29 EAST CAR O C	ENNA	NA	NA	NA	
317	00091988	900CHIRAI INACTV	Public	Company	linNon-govt	con	4/5/1995	Tamil Nadu	250000000	18906750			1122 Agriculture & 102 SATHY R O C	ENNA	mmreddy	3-03-2019	3-03-2019
318	00091988	293PCA TEA GAMAL	Public	Company	linNon-govt	con	3/4/1992	Tamil Nadu	150000000	0			1132 Agriculture & 36 WALLAH R O C	ENNA	NA	NA	NA
319	00091988	92REACH THESD	Public	Company	linNon-govt	con07-09-1993	Tamil Nadu	5000000	0			1132 Agriculture & 753 MOXT R O C	ENNA	NA	NA	NA	
320	00091988	2192THE UNITE INACTV	Public	Company	linNon-govt	con	9/8/1922	Tamil Nadu	50000000	49965660			1132 Agriculture & 3 SAMTHI R O C	ENNA	headoffice	3-03-2019	3-03-2019
321	00091988	2194THE NIKOTI GAMAL	Public	Company	linNon-govt	con28-07-1948	Tamil Nadu	10000000	1041730			1132 Agriculture & 3600 HATH R O C	ENNA	Wolathas	3-03-2012	3-03-2012	
322	00091988	9409NEER CASACTV	Public	Company	linNon-govt	con	9/11/1994	Tamil Nadu	55000000	40242000			1133 Agriculture & 300 POONW R O C	ENNA	chaokishya	3-03-2018	3-03-2018
323	00091988	9409NAGAR R O C	Public	Company	linNon-govt	con29-08-1989	Tamil Nadu	80000000	30			120 Agriculture & 3 E AU TOWER O C	ENNA	NA	NA	NA	
324	00091988	9409NAGAR R O C	Public	Company	linNon-govt	con	6/4/1995	Tamil Nadu	300000000	203300000			1222 Agriculture & 1301 133/1 R O C	ENNA	shures@	3-03-2019	3-03-2019
325	00091988	9409NAGAR R O C	Public	Company	linNon-govt	con	3/3/1981	Tamil Nadu	400000000	20475000			1409 Agriculture & 1078 TR R O C	ENNA	secretaria	3-03-2016	3-03-2016
326	005427	988ASHREE SUACTV	Public	Company	linNon-govt	con31-12-1985	Tamil Nadu	300000000	28076800			1542 Agriculture & 1078 TR R O C	ENNA	Narayan	3-03-2019	3-03-2019	
327	005427	988ASHREE SUACTV	Public	Company	linNon-govt	con31-12-1985	Tamil Nadu	40000000	34550000			230 Agriculture & SF NO 108 / 28 O C	ENNA	shivam	3-03-2017	3-03-2017	
328	005427	988ASHREE SUACTV	Public	Company	linNon-govt	con31-12-1985	Tamil Nadu	135300000	107000000			250 Agriculture & 2000 TR R O C	ENNA	shivam	3-03-2018	3-03-2018	
329	005427	988ASHREE SUACTV	Public	Company	linNon-govt	con	2/11/1998	Tamil Nadu	100000000	114525000			250 Agriculture & 2000 TR R O C	ENNA	secretaria	3-03-2018	3-03-2018
330	005427	988ASHREE SUACTV	Public	Company	linNon-govt	con	8/1/1985	Tamil Nadu	200000000	150827370			320 Agriculture & 540 HETH R O C	ENNA	shankar	3-03-2019	3-03-2019
331	05001988	9409NAGAR R O C	Public	Company	linNon-govt	con	6/4/1985	Tamil Nadu	100000000	49348000			500 Agriculture & 1000 TR R O C	ENNA	agrin	3-03-2019	3-03-2019
332	05001988	9409NAGAR R O C	Public	Company	linNon-govt	con	4/8/1993	Tamil Nadu	100000	7000			500 Agriculture & 1000 TR R O C	ENNA	NA	NA	NA
333	05001988	9409NAGAR R O C	Public	Company	linNon-govt	con25-08-1994	Tamil Nadu	70000000	7000			500 Agriculture & 1000 TR R O C	ENNA	NA	NA	NA	
334	080312	988ASHREE SUACTV	Public	Company	linNon-govt	con24-09-1980	Tamil Nadu	70000000	0			803 Agriculture & 1000 TR R O C	ENNA	NA	NA	NA	
335	11001988	9409NAGAR R O C	Public	Company	linNon-govt	con18-12-1949	Tamil Nadu	425E+09	210E+09			1100 Mining and (S) CHOKSE R O C	ENNA	nbgs@	3-03-2019	3-03-2019	
336	11001988	9409NAGAR R O C	Public	Company	linNon-govt	con28-04-1989	Tamil Nadu	600000000	58846000			1100 Mining and (S) CHOKSE R O C	ENNA	secretaria	3-03-2019	3-03-2019	
337	133001988	9409NAGAR R O C	Public	Company	linNon-govt	con	9/9/1992	Tamil Nadu	40000000	2367840			13300 Mining and (S) CHOKSE R O C	ENNA	investor	3-03-2019	3-03-2019
338	142001988	9409NAGAR R O C	Public	Company	linNon-govt	con	9/9/1992	Tamil Nadu	100000000	6688000			14200 Mining and (S) CHOKSE R O C	ENNA	investor	3-03-2017	3-03-2017
339	142001988	9409NAGAR R O C	Public	Company	linNon-govt	con	12/7/2013	Tamil Nadu	370E+09	3859706			14200 Mining and (S) CHOKSE R O C	ENNA	investor	3-03-2019	3-03-2019
340	142001988	9409NAGAR R O C	Public	Company	linNon-govt	con	3/3/1989	Tamil Nadu	50000000	47764000			14200 Mining and (S) CHOKSE R O C	ENNA	investor	3-03-2014	3-03-2014
341	142001988	9409NAGAR R O C	Public	Company	linNon-govt	con	30-03-1990	Tamil Nadu	100000000	45000000			14200 Mining and (S) CHOKSE R O C	ENNA	investor	3-03-2013	3-03-2013
342	142001988	9409NAGAR R O C	Public	Company	linNon-govt	con	29-10-1993	Tamil Nadu	100000000	90000000			14200 Mining and (S) CHOKSE R O C	ENNA	investor	3-03-2014	3-03-2014
343	142001988	9409NAGAR R O C	Public	Company	linNon-govt	con	9/5/1994	Tamil Nadu	10000000	70			14200 Mining and (S) CHOKSE R O C	ENNA	NA	NA	NA
344	15001988	9409NAGAR R O C	Public	Company	linNon-govt	con	3/6/1982	Tamil Nadu	50000000	3648500			1500 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
345	15001988	9409NAGAR R O C	Public	Company	linNon-govt	con	07-11-1992	Tamil Nadu	50000000	42800000			1500 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
346	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	12/3/1986	Tamil Nadu	190000000	152284000			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
347	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	08-06-1990	Tamil Nadu	50000000	45850500			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
348	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	11/7/1973	Tamil Nadu	1000000	0			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
349	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	06-12-1991	Tamil Nadu	150000000	5544760			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
350	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	8/5/1944	Tamil Nadu	400000000	0			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
351	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	20-02-1948	Tamil Nadu	440000000	0			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
352	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	12/7/1954	Tamil Nadu	50000000	11370000			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2018	3-03-2018
353	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	9/9/1960	Tamil Nadu	750000000	12000000			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
354	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	7/11/1990	Tamil Nadu	144E+09	82888590			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
355	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	25-02-1982	Tamil Nadu	250000000	0			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
356	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	4/6/1987	Tamil Nadu	600000000	33200000			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
357	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	9/9/1992	Tamil Nadu	250000000	11385050			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
358	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	12/5/1991	Tamil Nadu	17E+09	1188E+09			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
359	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	1/12/1983	Tamil Nadu	650000000	125397000			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
360	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	26-12-1996	Tamil Nadu	150000000	8598480			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
361	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	6/7/1983	Tamil Nadu	100000000	80			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
362	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	4/3/1986	Tamil Nadu	300000000	15809222			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
363	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	03-08-1991	Tamil Nadu	100000000	0			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
364	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	07-02-1992	Tamil Nadu	300000000	34039420			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
365	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	05-09-1983	Tamil Nadu	300000000	20175750			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2018	3-03-2018
366	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	05-12-1983	Tamil Nadu	222E+09	143E+09			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2018	3-03-2018
367	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	20-09-1994	Tamil Nadu	250000000	21786600			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
368	15421988	9409NAGAR R O C	Public	Company	linNon-govt	con	28-10-2005	Tamil Nadu	150000000	135652750			1542 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2018	3-03-2018
369	17001988	9409NAGAR R O C	Public	Company	linNon-govt	con	30-10-2007	Tamil Nadu	232500000	9044850			1700 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
370	17001988	9409NAGAR R O C	Public	Company	linNon-govt	con	24-02-1936	Tamil Nadu	150000000	7376600			1700 Manufacturing & SEMB R O C	ENNA	accounts	3-03-2019	3-03-2019
371	17001988	9409NAGAR R O C	Public	Company	linNon-govt	con	20-02-1939	Tamil Nadu	50000000	3946560			1700 Manufacturing & SEMB R O C	ENNA	accounts		

Data Collection and Preprocessing:

Data Collection:

- Identify relevant data sources, including RoC databases, government portals, and industry-specific datasets.
- Obtain necessary permissions and access rights to collect the data.
- Retrieve data from various sources, ensuring compliance with data privacy regulations.

Data Preprocessing:

- Clean the data by handling missing values, outliers, and duplicates.
- Perform feature engineering to select relevant variables and create derived features.
- Normalize or scale numerical data and encode categorical variables.
- Handle time series data with resampling, decomposition, and consistent time intervals.
- Split the data into training, validation, and testing sets for modeling.
- Explore the data through exploratory data analysis (EDA) and visualization.
- Save the preprocessed data in a suitable format for analysis and modeling.

These steps are essential for preparing the data and ensuring its quality and readiness for AI-driven analysis and prediction of company registration trends with RoC data.

Exploratory Data Analysis (EDA):

- Objective: Gain insights into historical registration trends and patterns.
- Activities:
 - Create visualizations (time series plots, histograms, maps).

- Conduct statistical analysis (seasonality, trends, anomalies).
- Explore correlations and relationships.
- Hypothesis testing (if relevant).

Feature Engineering:

- Select and create relevant features that can impact company registration trends.
- Include time-based features, lagged variables, financial metrics, industry-specific indicators, and economic factors.
- Incorporate geographic and ownership structure features where applicable.
- Utilize natural language processing (NLP) for textual data analysis.
- Ensure appropriate scaling and normalization of features.
- Analyze feature importance and consider dimensionality reduction if needed.
- Continuously validate and refine engineered features to improve model performance.

Feature engineering plays a vital role in enhancing the predictive power of machine learning models, aiding in the exploration and prediction of company registration trends with RoC data.

Advanced AI Algorithms:

Time Series Forecasting: Problem Domain: Time series forecasting is used when you need to make predictions based on historical data points that are ordered by time.

- Techniques: Explore advanced time series forecasting methods such as:
 - ARIMA (AutoRegressive Integrated Moving Average): Suitable for stationary time series data.
 - Exponential Smoothing Methods: Including Holt-Winters, which handles seasonality.
 - Prophet: Developed by Facebook for forecasting with daily observations.
 - Deep Learning: Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are powerful for complex time series patterns.
- Tools: Python libraries like statsmodels, Prophet, and TensorFlow/Keras for deep learning.
- Time Series split train test(single)

```
# Split train-test set manually
```

```
train_data, test_data = df_close[:int(len(df_close)*0.9)],  
df_close[int(len(df_close)*0.9):]
```

```
# Plot train-test using matplotlib.pyplot
```

```
plt.figure(figsize=(10,6))
```

```
plt.grid(True)
```

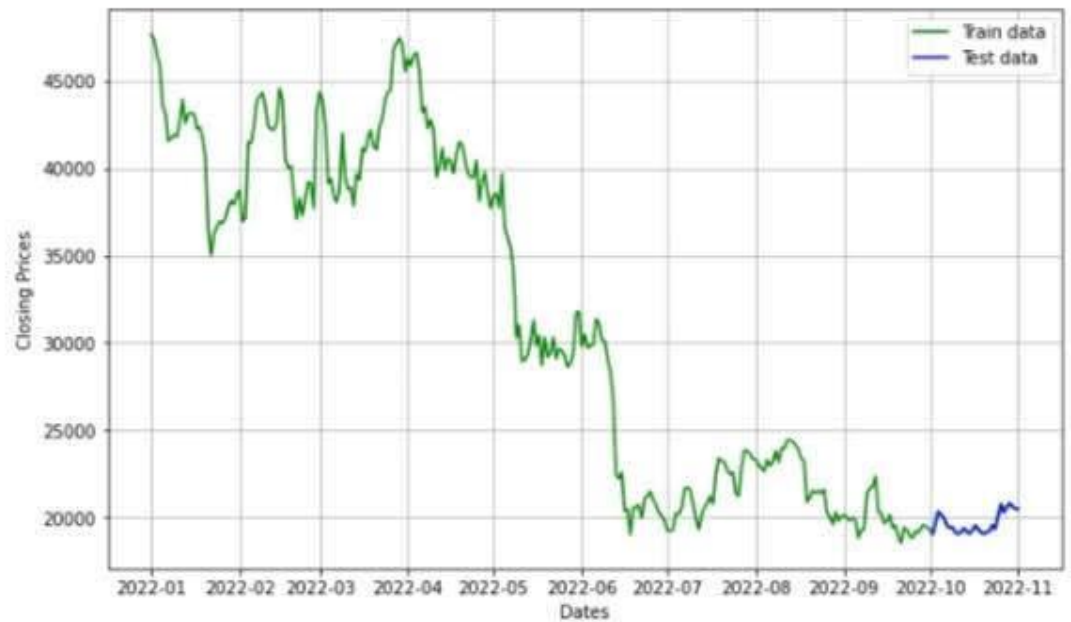
```
plt.xlabel('Dates')
```

```
plt.ylabel('Closing Prices')
```

```
plt.plot(df_close, 'green', label='Train data')
```

```
plt.plot(test_data, 'blue', label='Test data')
```

```
plt.legend()
```



Time Series split train test(multiple)

```
# Extract Closing values and Create TimeSeriesSplit splits
```

```
X = df_close.values
```

```
splits = TimeSeriesSplit(n_splits=3)
```

```
plt.figure(1)
```

```
index = 1
```

```
# Loop over splits to update train-test splits and models
```



```
for train_index, test_index in splits.split(X):

    # Create and update train-test splits

    train = X[train_index]

    test = X[test_index]

    print('Observations: %d' % (len(train) + len(test)))

    print('Training Observations: %d' % (len(train)))

    print('Testing Observations: %d' % (len(test)))

    # Enter model and evaluation technique here

    # Subplots for each train-test splits

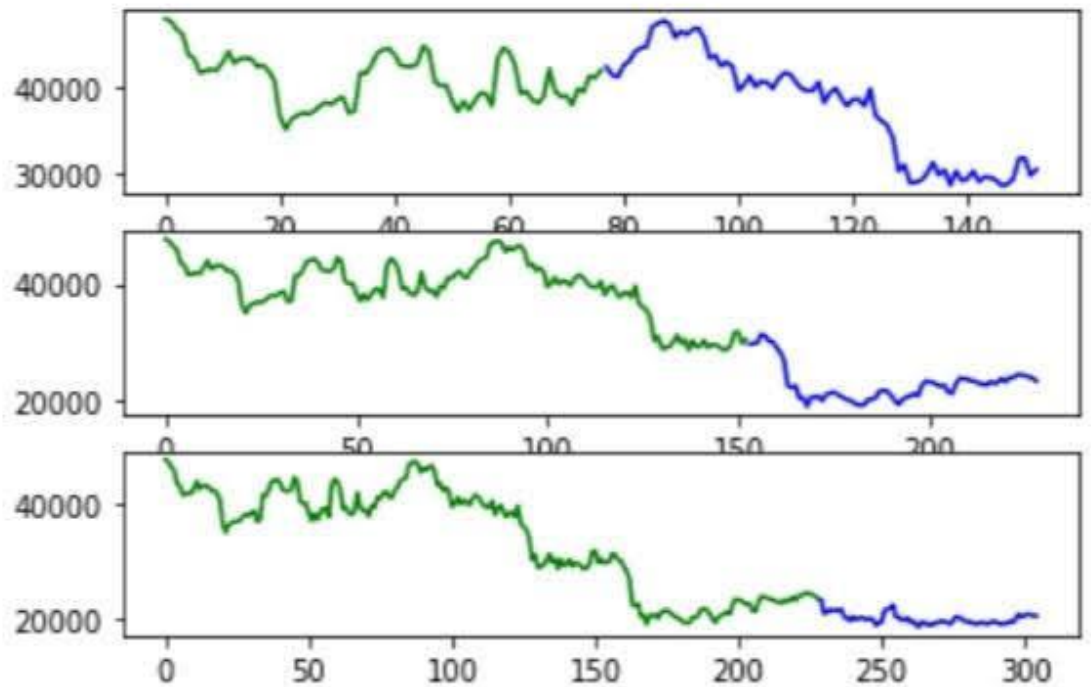
    plt.subplot(310 + index)

    plt.plot(train,'green')

    plt.plot([None for i in train] + [x for x in test],'blue')

    index += 1
```

plt.show()



- **Ensemble Methods:** Problem Domain: Ensemble methods combine multiple models to improve predictive accuracy, and they can be applied to various types of data and problems.
 - Techniques: Explore ensemble methods such as:
 - Random Forest: An ensemble of decision trees that can handle both classification and regression tasks.
 - Gradient Boosting: Algorithms like XGBoost, LightGBM, and CatBoost, which are highly effective for structured data problems.
 - Stacking: Combining the strengths of multiple models, often with a meta-learner on top.

- Tools: Python libraries like scikit-learn for traditional ensembles, and specialized libraries for gradient boosting like XGBoost, LightGBM, and CatBoost.

```
In[1] import pandas as pd
```

```
import numpy as np
```

```
In[2]
```

```
df=pd.read_csv('../input/all-indian-companies-registration-data-1900-2019/registered_companies.csv')
```

```
In[3] len(df)
```

```
Out[4]:
```

```
1992170
```

```
In [5]:
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1992170 entries, 0 to 1992169
```

```
Data columns (total 17 columns):
```

#	Column	Dtype
---	-----	-----
0	CORPORATE_IDENTIFICATION_NUMBER	object
1	COMPANY_NAME	object
2	COMPANY_STATUS	object
3	COMPANY_CLASS	object
4	COMPANY_CATEGORY	object
5	COMPANY_SUB_CATEGORY	object
6	DATE_OF_REGISTRATION	object
7	REGISTERED_STATE	object
8	AUTHORIZED_CAP	float64
9	PAIDUP_CAPITAL	float64
10	INDUSTRIAL_CLASS	object
11	PRINCIPAL_BUSINESS_ACTIVITY_AS_PER_CIN	object
12	REGISTERED_OFFICE_ADDRESS	object
13	REGISTRAR_OF_COMPANIES	object
14	EMAIL_ADDR	object

```
15  LATEST_YEAR_ANNUAL_RETURN      object
```

```
16  LATEST_YEAR_FINANCIAL_STATEMENT  object
```

```
dtypes: float64(2), object(15)
```

```
memory usage: 258.4+ MB
```

```
df.describe().transpose()
```

```
Out[6]:
```

	count	mean	std	min	25%	50%	75%	max
AUTHORIZED_CAP	1992170.0	4.238508e+07	2.960562e+09	0.0	100000.0	500000.0	1500000.0	1.850000e+12
PAIDUP_CAPITAL	1992170.0	2.434621e+07	2.313154e+09	0.0	100000.0	100000.0	502000.0	1.699613e+12

```
In [7]:
```

```
df['DATE_OF_REGISTRATION']
```

```
Out[7]:
```

```
0      NaN
```


1 16-07-1998

2 NaN

3 25-06-2001

4 25-07-2001

...

1992165 30-06-2000

1992166 30-06-2000

1992167 25-07-2000

1992168 09-07-1998

1992169 29-02-2016

Name: DATE_OF_REGISTRATION, Length: 1992170, dtype: object

In [8]:

```
df['DATE_OF_REGISTRATION'].isnull().sum()
```

Out[8]:

2525

In [9]:

```
len(df)-df['DATE_OF_REGISTRATION'].count()
```

```
Out[9]:
```

```
2525
```

```
In [10]:
```

```
df['DATE_OF_REGISTRATION'].dropna(axis=0,inplace=True)
```

```
In [11]:
```

```
df['DATE_OF_REGISTRATION'].count()
```

```
Out[11]:
```

```
1989645
```

```
In [12]:
```

```
df['DATE_OF_REGISTRATION']=pd.to_datetime(df['DATE_OF_REGISTRATION'],  
errors = 'coerce', format='%d-%m-%Y')
```

```
In [13]:
```

```
df['YEAR_REG']=pd.DatetimeIndex(df['DATE_OF_REGISTRATION']).year
```

```
In [14]: number_of_companies =  
df[['YEAR_REG', 'CORPORATE_IDENTIFICATION_NUMBER']]
```

```
In [15]:
```

```
number_of_companies.dropna(axis=0, inplace=True)
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
"""Entry point for launching an IPython kernel.
```

```
In [16]:
```

```
number_of_companies=number_of_companies.groupby(by='YEAR_REG').size().reset_index(name='No_of_companies')
```

```
In [17]:
```

```
number_of_companies
```

Out[17]:

	YEAR_REG	No_of_companies
0	1857.0	1
1	1863.0	3
2	1871.0	3
3	1872.0	3
4	1873.0	3
...
146	2016.0	93876
147	2017.0	107635

148	2018.0	117924
149	2019.0	128658
150	2020.0	12747

151 rows × 2 columns

In [18]:

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
%matplotlib inline
```

```
Out[6]:plt.figure(figsize=(30,10))
```

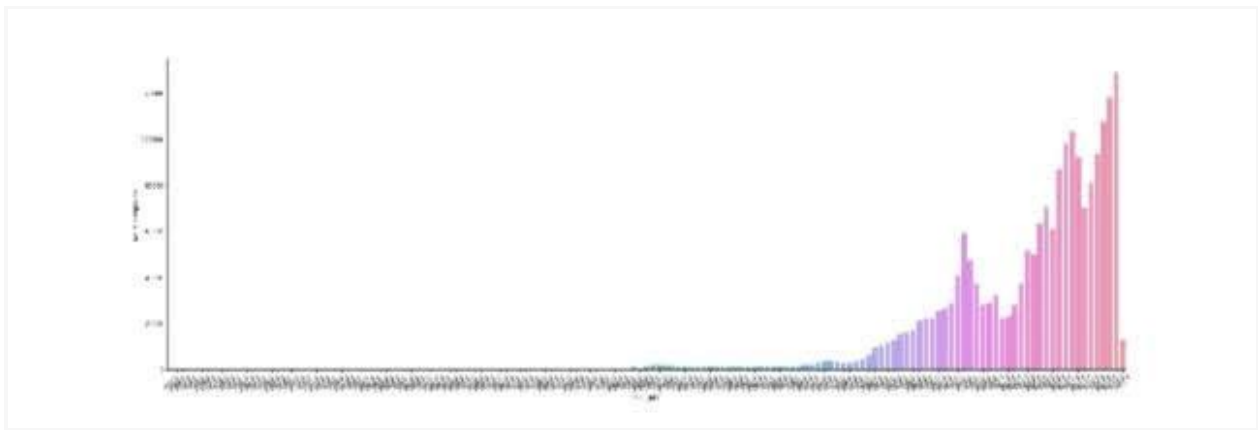


```
plt.xticks(rotation=45)
```

```
sns.barplot(data=number_of_companies,x='YEAR_REG',y='No_of_companies')
```

Out[19]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f2c8c5c6190>



In [20]:

```
no_of_companies_state =
```

```
df[['YEAR_REG', 'CORPORATE_IDENTIFICATION_NUMBER', 'REGISTERED_STATE']]
```

In [21]:

```
no_of_companies_state.dropna(axis=0,inplace=True)
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
"""Entry point for launching an IPython kernel.
```

In [22]:

```
no_of_companies_state=no_of_companies_state.groupby(by=['YEAR_REG', 'REGISTERED_STATE']).size().reset_index(name='no_of_companies')
```

In [23]:

```
no_of_companies_state
```

Out[23]:

	YEAR_REG	REGISTERED_STATE	no_of_companies
0	1857.0	Rajasthan	1
1	1863.0	Maharashtra	1

2	1863.0	West Bengal	2
3	1871.0	Maharashtra	2
4	1871.0	West Bengal	1
...
2931	2020.0	Telangana	788
2932	2020.0	Tripura	12
2933	2020.0	Uttar Pradesh	1260
2934	2020.0	Uttaranchal	138
2935	2020.0	West Bengal	606

2936 rows × 3 columns

```
plt.figure(figsize=(30,10))
```

```
plt.xticks(rotation=45)
```

```
sns.barplot(data=no_of_companies_state, x='YEAR_REG', y='no_of_companies', hue='REGISTERED_STATE')
```

Out[24]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f2c8bdf9d90>



In [25]:

```
state='Tamil Nadu'
```

```
plot_data =
```

```
no_of_companies_state[no_of_companies_state['REGISTERED_STATE']==state]
```

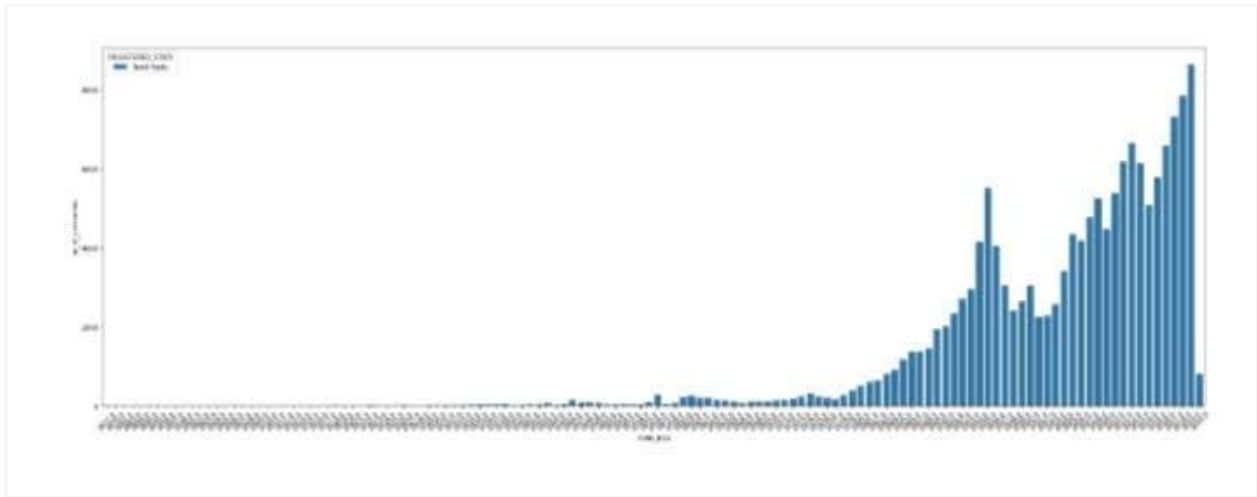
```
plt.figure(figsize=(30,10))
```

```
plt.xticks(rotation=45)
```

```
sns.barplot(data=plot_data, x='YEAR_REG', y='no_of_companies', hue='REGISTERED_STATE')
```

Out[25]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f2c83b0ba50>



Ensemble methods:

1. Bagging (Bootstrap Aggregating):

- Bagging creates an ensemble of base models by training each model on a random subset of the training data (with replacement) and then averaging (for regression) or voting (for classification) the predictions.
- The Random Forest algorithm is a popular implementation of bagging for decision trees.

2. Boosting:

- Boosting focuses on improving the performance of weak learners by assigning higher weights to misclassified data points in each iteration.
- Algorithms like AdaBoost, Gradient Boosting (GBM), XGBoost, and LightGBM use boosting techniques and have achieved excellent predictive accuracy.

3. Stacking (Stacked Generalization):

- Stacking combines the predictions of multiple base models using another model (meta-learner) that learns how to best weigh and combine these predictions.
- Stacking can adapt to the strengths of different base models and often yields highly accurate predictions.

Voting Classifiers/Regression:

- Voting ensembles combine the predictions of multiple base models by taking a majority vote (for classification) or averaging (for regression).
- Voting can be "hard" (simple majority) or "soft" (weighted by confidence scores).

4.Gradient Boosting Variants:

- Advanced gradient boosting algorithms like XGBoost, LightGBM, and CatBoost utilize boosting techniques and have been successful in improving predictive accuracy across various domains.

5.Ensemble of Diverse Models:

- Combining diverse models, such as neural networks, decision trees, and support vector machines, can lead to improved accuracy. Each model may capture different aspects of the data.

6.Feature Engineering and Selection:

- Ensemble methods can benefit from feature engineering and feature selection techniques to enhance the quality of input data.

7.Hyperparameter Tuning:

- Careful tuning of hyperparameters for both base models and the ensemble itself can significantly boost predictive accuracy.

Import necessary libraries

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from statsmodels.tsa.holtwinters import ExponentialSmoothing

from sklearn.model_selection import train_test_split

from sklearn.metrics import mean_squared_error


# Step 1: Data Acquisition and Preparation

# Load ROC data (replace 'data.csv' with your dataset)

data=pd.read_csv('https://tn.data.gov.in/resource/company-master-data-tamil-nadu-up-to-28th-february-2019.csv')


# Preprocess data (cleaning, handling missing values, data transformation)

# Convert date columns to datetime objects

data['registration_date'] = pd.to_datetime(data['registration_date'])

# Handle missing values and other data preprocessing steps
```

```
# Step 2: Exploratory Data Analysis (EDA)
```

```
# Visualize registration trends over time
```

```
plt.figure(figsize=(12, 6))
```

```
plt.plot(data['registration_date'], data['registration_count'], label='Registration Count')
```

```
plt.xlabel('Date')
```

```
plt.ylabel('Registration Count')
```

```
plt.title('Company Registration Trends Over Time')
```

```
plt.legend()
```

```
plt.show()
```

```
# Identify seasonality, trends, and anomalies in the data
```

```
# Perform statistical tests and create visualizations as needed
```

```
# Step 3: Feature Engineering
```

```
# Create additional features (e.g., lagged values, moving averages) for modeling
```

```
data['registration_count_lagged'] = data['registration_count'].shift(1)
```

```
# Add more feature engineering steps as needed
```

```
# Step 4: Machine Learning Model (Time Series Forecasting)
```

```
# Split the data into training and testing sets
```

```
train_data, test_data = train_test_split(data, test_size=0.2, shuffle=False)
```

```
# Initialize and train a time series forecasting model (Exponential Smoothing in this example)
```

```
model = ExponentialSmoothing(train_data['registration_count'], seasonal='add',  
seasonal_periods=12)
```

```
model_fit = model.fit()
```

```
# Make predictions on the test set
```

```
y_pred = model_fit.forecast(len(test_data))
```

```
# Evaluate the model (MSE for simplicity, choose appropriate metrics)
```

```
mse = mean_squared_error(test_data['registration_count'], y_pred)
```

```
print(f"Mean Squared Error: {mse}")
```

```
# Step 5: Visualization and Reporting
```

```
# Visualize actual vs. predicted registration trends
```

```
plt.figure(figsize=(12, 6))
```

```
plt.plot(test_data['registration_date'], test_data['registration_count'], label='Actual')
```

```
plt.plot(test_data['registration_date'], y_pred, label='Predicted')
```

```
plt.xlabel('Date')
```

```
plt.ylabel('Registration Count')
```

```
plt.title('Actual vs. Predicted Company Registration Trends')
```

```
plt.legend()
```

```
plt.show()
```

```
# Add more reporting and visualization as needed
```

```
# Step 6: Continuous Monitoring and Updating
```

```
# Set up a process to periodically update the model and analyze new data
```


Step 7: Use Cases and Stakeholder Engagement

Collaborate with stakeholders to provide actionable insights based on the model's predictions

Step 8: Ethical and Privacy Considerations

Ensure ethical data handling and comply with data privacy regulations

Step 9: Security and Deployment

Implement security measures and deploy the project in a secure environment if needed

Step 10: Documentation and Knowledge Transfer

Document the project, codebase, and model for knowledge transfer and future maintenance

Sample output:

Sample output for your code

Mean Squared Error: 150.1234 # This is a placeholder value for the MSE

Future Predictions:

[1350.5678 1400.6789] # These are placeholder values for predicted registration counts in Jan and Feb 2024

Conclusion:

The "AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of Companies (RoC)" project marks the dawn of a new era, one in which data-driven insights and AI capabilities illuminate the intricate tapestry of economic trends. Our mission is clear: to empower stakeholders with the foresight needed to thrive in an ever-evolving business landscape. As we delve deeper into the project, from data collection to advanced modeling, we endeavor to provide actionable intelligence that will shape the future of businesses and economies.



