SURGE 2024 - All Codes

- **Converting Coordinates to City Names**

```python
from geopy.geocoders import Nominatim
import pandas as pd
from tqdm import tqdm

tqdm.pandas()

data_2022 = pd.read_csv('/content/India2022Location.csv',
low_memory=False)
data_2014 = pd.read_csv('/content/India2014Location.csv',
low_memory=False)

geolocator = Nominatim(timeout=10, user_agent = "abcd")

data_2022['Location_Coordinate'] = data_2022.apply(lambda row:
f"{row['lat_mask']},{row['lon_mask']}", axis=1)
data_2014['Location_Coordinate'] = data_2014.apply(lambda row:
f"{row['lat_mask']},{row['lon_mask']}", axis=1)

def get_district(x):
    try:
        location = geolocator.reverse(x)
        return location.raw['address'].get('state_district', '-1')
    except Exception as e:
        # print(f"Error processing {x}: {e}")
        return '-1'
```

geolocator.reverse(x).raw['address'] can have either city or county extract
which ever is available
```python
data_2014['District'] =
data_2014['Location_Coordinate'].progress_apply(lambda x:
```

```python
geolocator.reverse(x).raw['address']['state_district'] if 'state_district' in
geolocator.reverse(x).raw['address'] else '-1')
data_2014['District'] =
data_2014['Location_Coordinate'].progress_apply(get_district)

data_2014.to_csv('India2014LocationUpdated.csv', index=False)
data_2014.to_csv('India2014LocationUpdated.csv', index=False)
```

- **Processing Delhi - Meerut Expressway Data & Running DiD**

```r
rm(list = ls())

directory <- "/Users/derpboy77/Desktop/SURGE"
setwd(directory)

install.packages("nnet")
library(nnet)

library(rio)
library(dplyr)
library(tidyr)
library(stargazer)

df_2022 <- rio::import("data_2022.xlsx")
df_2014 <- rio::import("data_2014.xlsx")

df_2022 <- df_2022 %>% select(idstd, Treatment, Post, d2, d30a, l2)
df_2014 <- df_2014 %>% select(idstd, Treatment, Post, d2, d30a, l2)

df_2022 <- drop_na(df_2022)
df_2014 <- drop_na(df_2014)

table(df_full$l2)

df_full <- bind_rows(df_2022, df_2014)
```

```
df_full <- df_full %>% filter(d2 != -9)
df_full <- df_full %>% filter(l2 != -9)
df_full <- df_full %>% filter(l2 != -7)

df_full$d2 <- (df_full$d2 - mean(df_full$d2))/sd(df_full$d2)

# DiD model for sales
gm_1 <- lm(d2~Treatment+Post+Treatment*Post+l2, df_full)
stargazer(gm_1, type="text")

table(df_full$d30a)
df_full <- df_full %>% filter(d30a != -9)
df_full <- df_full %>% filter(d30a != -7)

# OLS DiD on perception of transport as an obstacle
gm_2 <- lm(d30a~Treatment+Post+Treatment*Post, df_full)
stargazer(gm_2, type="text")

# Multinomial Logistic Regression on perception of transport as an
obstacle
gm_3 <- multinom(d30a~Treatment+Post+Treatment*Post, df_full)
stargazer(gm_3, type="text")
```

● **Implementing Saugato Datta's Paper**

```
rm(list = ls())

directory <- "/Users/derpboy77/Desktop/SURGE"
setwd(directory)

library(rio)
library(dplyr)
library(tidyr)
library(stargazer)
```

```r
data<-rio::import("India-2005--full-data-.dta")
data_2002<-rio::import("India2002_M_clean.dta")
a <- rio::import("India-2022-full-data.dta")

onGQ_cities <- c(1,2,3,5,7,20,36,15,19,34,22,9,10,37,11,17,30,18,13)
offGQ_cities <- c(25,23,4,6,33,26,8,27,31,14,32,35,40,12,28,29,24,21)
nodal_cities <- c(7,10,3,5,19,20,36,37,30)

considered_cities = c(onGQ_cities, offGQ_cities)

df <- data %>% filter(code3 %in% considered_cities)
df_2002 <- data_2002 %>% filter(code3 %in% considered_cities)

df <- df %>%
  mutate(
    onGQ = case_when(
      code3 %in% onGQ_cities ~ 1,
      TRUE ~ 0
    ),
    offGQ = case_when(
      code3 %in% offGQ_cities ~ 1,
      TRUE ~ 0
    ),
    Distance_from_GQ = case_when(
      code3 == 25 ~ 541,
      code3 == 23 ~ 355,
      code3 == 4 ~ 238,
      code3 == 6 ~ 533,
      code3 == 33 ~ 340,
      code3 == 25 ~ 541,
      code3 == 26 ~ 118,
      code3 == 8 ~ 267,
      code3 == 27 ~ 415,
      code3 == 31 ~ 364,
      code3 == 14 ~ 77,
      code3 == 32 ~ 305,
      code3 == 35 ~ 444,
```

```
      code3 == 40 ~ 347,
      code3 == 12 ~ 139,
      code3 == 28 ~ 747,
      code3 == 29 ~ 185,
      code3 == 24 ~ 392,
      code3 == 21 ~ 85,
      TRUE ~ 0
    )
  )

df_2002 <- df_2002 %>%
  mutate(
    onGQ = case_when(
      code3 %in% onGQ_cities ~ 1,
      TRUE ~ 0
    ),
    offGQ = case_when(
      code3 %in% offGQ_cities ~ 1,
      TRUE ~ 0
    ),
    Distance_from_GQ = case_when(
      code3 == 25 ~ 541,
      code3 == 23 ~ 355,
      code3 == 4 ~ 238,
      code3 == 6 ~ 533,
      code3 == 33 ~ 340,
      code3 == 25 ~ 541,
      code3 == 26 ~ 118,
      code3 == 8 ~ 267,
      code3 == 27 ~ 415,
      code3 == 31 ~ 364,
      code3 == 14 ~ 77,
      code3 == 32 ~ 305,
      code3 == 35 ~ 444,
      code3 == 40 ~ 347,
      code3 == 12 ~ 139,
      code3 == 28 ~ 747,
```

```r
      code3 == 29 ~ 185,
      code3 == 24 ~ 392,
      code3 == 21 ~ 85,
      TRUE ~ 0
    )
  )

industries <- 1:15

df <- filter(df, code2 %in% industries)


df %>% filter(code2 == 15) %>% nrow()


df_fil <- df %>% select(code1, code2, code3, r3_3, r3_1a1, r11_5ac,onGQ,
offGQ, Distance_from_GQ)
df_fil <- drop_na(df_fil)

df_fil_2002 <- df_2002 %>% select(code1, code2, code3, q3191, q3121,
q907c, onGQ, offGQ, Distance_from_GQ)
df_fil_2002 <- drop_na(df_fil_2002)

df_fil <- df_fil %>% mutate(Post = 1)
df_fil_2002 <- df_fil_2002 %>% mutate(Post = 0)

df_fil_2002 <- df_fil_2002 %>% rename(r3_3 = q3191)
df_fil_2002 <- df_fil_2002 %>% rename(r3_1a1 = q3121)
df_fil_2002 <- df_fil_2002 %>% rename(r11_5ac = q907c)

mean(df_fil[df_fil$onGQ == 1, "r3_3"])
mean(df_fil_2002[df_fil_2002$onGQ == 1, "r3_3"])

mean(df_fil[df_fil$offGQ == 1, "r3_3"])
mean(df_fil_2002[df_fil_2002$offGQ == 1, "r3_3"])

mean(df_fil[df_fil$onGQ == 1, "r3_1a1"])
```

```r
mean(df_fil_2002[df_fil_2002$onGQ == 1, "r3_1a1"])

mean(df_fil[df_fil$offGQ == 1, "r3_1a1"])
mean(df_fil_2002[df_fil_2002$offGQ == 1, "r3_1a1"])


df_full <- bind_rows(df_fil, df_fil_2002)
df_full_non_nodal <- df_full %>% filter(!(code3 %in% nodal_cities))

gm1 <- lm(r3_3~onGQ+Post+onGQ*Post, df_full)
gm1_1 <- lm(r3_3~onGQ+Post+onGQ*Post, df_full_non_nodal)

stargazer(gm1, gm1_1, type="text")

gm2 <- lm(r3_1a1~onGQ+Post+onGQ*Post, df_full)
gm2_1 <- lm(r3_1a1~onGQ+Post+onGQ*Post, df_full_non_nodal)
stargazer(gm2, gm2_1, type="text")


gm3 <- lm(r11_5ac~onGQ+Post+onGQ*Post, df_full)
gm3_1 <- lm(r11_5ac~onGQ+Post+onGQ*Post, df_full_non_nodal)
stargazer(gm3, gm3_1, type="text")

mean(df_full_non_nodal[df_full_non_nodal$Post == 1 &
df_full_non_nodal$onGQ == 1, "r11_5ac"])

install.packages("Hmisc")
library(Hmisc)

labels_df <- data.frame(
  Column = names(a),
  Label = sapply(a, function(x) label(x))
)
print(labels_df)
```

r3_1a1 / q3121 - No. of years in business with main input supplier
r3_3 / q3191 - No. of years in business with input
r11_5ac / q907c - transportation as obstacle