Creating Shape files for Varanasi
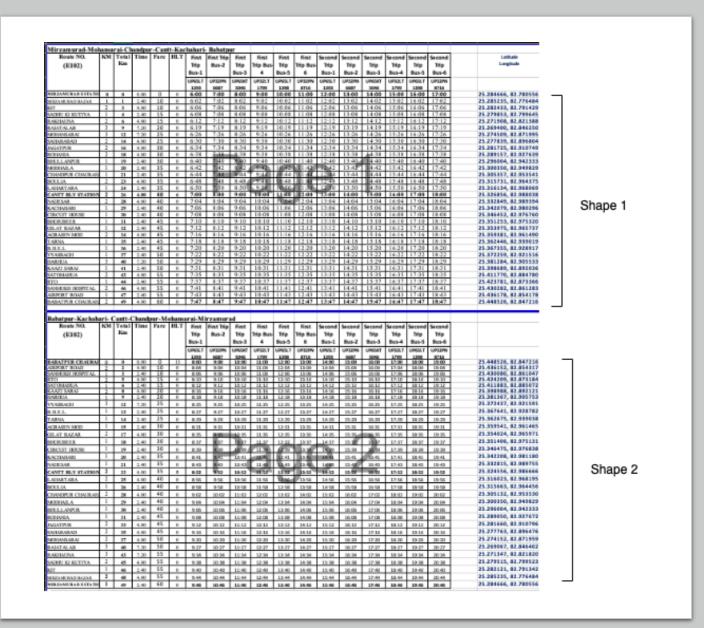
Ammar Bin Zulqarnain

Example Route Sheet

Mirzamurad-Mohansarai-Chandpur-Cantt-Kachahari- Babatpur

Route NO. (E102)	KM	Total Km	Time	Fare	HLT	First Trip Bus-1 UP65LT 1393	First Trip Bus-2 UP32PN 6687	First Trip Bus-3 UP65KT 5046	First Trip Bus 4 UP32LT 1799	First Trip Bus-5 UP65LT 1398	First Trip Bus 6 UP32PN 8716	Second Trip Bus-1 UP65LT 1393	Second Trip Bus-2 UP32PN 6687	Second Trip Bus-3 UP65KT 5046	Second Trip Bus-4 UP32LT 1799	Second Trip Bus-5 UP65LT 1398	Second Trip Bus-6 UP32PN 8716	Latitude Longitude
MIRZAMURAD ESTATIO	0	0	0.00	0	0	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	25.284666, 82.780556
MIRZAMURAD BAZAR	1	1	2.40	10	0	6:02	7:02	8:02	9:02	10:02	11:02	12:02	13:02	14:02	15:02	16:02	17:02	25.285235, 82.776484
KIT	2	3	4.80	10	0	6:06	7:06	8:06	9:06	10:06	11:06	12:06	13:06	14:06	15:06	16:06	17:06	25.282433, 82.791429
SADHU KI KUTIYA	1	4	2.40	15	0	6:08	7:08	8:08	9:08	10:08	11:08	12:08	13:08	14:08	15:08	16:08	17:08	25.279853, 82.799645
RAKHAUNA	2	6	4.80	15	0	6:12	7:12	8:12	9:12	10:12	11:12	12:12	13:12	14:12	15:12	16:12	17:12	25.271908, 82.821388
RAJATALAB	3	9	7.20	20	0	6:19	7:19	8:19	9:19	10:19	11:19	12:19	13:19	14:19	15:19	16:19	17:19	25.269400, 82.846250
MOHANSARAI	3	12	7.20	25	0	6:26	7:26	8:26	9:26	10:26	11:26	12:26	13:26	14:26	15:26	16:26	17:26	25.274509, 82.871995
SAHABABAD	2	14	4.80	25	0	6:30	7:30	8:30	9:30	10:30	11:30	12:30	13:30	14:30	15:30	16:30	17:30	25.277839, 82.896804
JAGATPUR	2	16	4.80	30	0	6:34	7:34	8:34	9:34	10:34	11:34	12:34	13:34	14:34	15:34	16:34	17:34	25.281725, 82.910749
ROHANIA	2	18	4.80	30	0	6:38	7:38	8:38	9:38	10:38	11:38	12:38	13:38	14:38	15:38	16:38	17:38	25.289157, 82.927639
BHULLANPUR	1	19	2.40	30	0	6:40	7:40	8:40	9:40	10:40	11:40	12:40	13:40	14:40	15:40	16:40	17:40	25.296004, 82.942333
MODHAILA	1	20	2.40	35	0	6:42	7:42	8:42	9:42	10:42	11:42	12:42	13:42	14:42	15:42	16:42	17:42	25.300350, 82.949829
CHANDPUR CHAURAH	1	21	2.40	35	0	6:44	7:44	8:44	9:44	10:44	11:44	12:44	13:44	14:44	15:44	16:44	17:44	25.305357, 82.953541
BOULIA	2	23	4.80	35	0	6:48	7:48	8:48	9:48	10:48	11:48	12:48	13:48	14:48	15:48	16:48	17:48	25.315731, 82.964375
LAHARTARA	1	24	2.40	35	0	6:50	7:50	8:50	9:50	10:50	11:50	12:50	13:50	14:50	15:50	16:50	17:50	25.316134, 82.968069
CANTT RLY STATION	2	26	4.80	40	6	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	25.326856, 82.988038
NADESAR	2	28	4.80	40	0	7:04	8:04	9:04	10:04	11:04	12:04	13:04	14:04	15:04	16:04	17:04	18:04	25.332849, 82.989394
KACHAHARI	1	29	2.40	40	0	7:06	8:06	9:06	10:06	11:06	12:06	13:06	14:06	15:06	16:06	17:06	18:06	25.342079, 82.980296
CIRCUIT HOUSE	1	30	2.40	40	0	7:08	8:08	9:08	10:08	11:08	12:08	13:08	14:08	15:08	16:08	17:08	18:08	25.346452, 82.976760
BHOJUBEER	1	31	2.40	45	0	7:10	8:10	9:10	10:10	11:10	12:10	13:10	14:10	15:10	16:10	17:10	18:10	25.351253, 82.975320
GILAT BAZAR	1	32	2.40	45	0	7:12	8:12	9:12	10:12	11:12	12:12	13:12	14:12	15:12	16:12	17:12	18:12	25.353975, 82.965737
AGRASEN MOD	2	34	4.80	45	0	7:16	8:16	9:16	10:16	11:16	12:16	13:16	14:16	15:16	16:16	17:16	18:16	25.359381, 82.961490
TARNA	1	35	2.40	45	0	7:18	8:18	9:18	10:18	11:18	12:18	13:18	14:18	15:18	16:18	17:18	18:18	25.362446, 82.939019
B.H.E.L	1	36	2.40	45	0	7:20	8:20	9:20	10:20	11:20	12:20	13:20	14:20	15:20	16:20	17:20	18:20	25.367355, 82.928917
VYASBAGH	1	37	2.40	50	0	7:22	8:22	9:22	10:22	11:22	12:22	13:22	14:22	15:22	16:22	17:22	18:22	25.372259, 82.921516
HARHUA	3	40	7.20	50	0	7:29	8:29	9:29	10:29	11:29	12:29	13:29	14:29	15:29	16:29	17:29	18:29	25.381284, 82.905533
KAAZI SARAI	1	41	2.40	50	0	7:31	8:31	9:31	10:31	11:31	12:31	13:31	14:31	15:31	16:31	17:31	18:31	25.398689, 82.892036
SATOMAHUA	2	43	4.80	55	0	7:35	8:35	9:35	10:35	11:35	12:35	13:35	14:35	15:35	16:35	17:35	18:35	25.411770, 82.884780
RTO	1	44	2.40	55	0	7:37	8:37	9:37	10:37	11:37	12:37	13:37	14:37	15:37	16:37	17:37	18:37	25.423781, 82.873366
SANMUKH HOSPITAL	2	46	4.80	55	0	7:41	8:41	9:41	10:41	11:41	12:41	13:41	14:41	15:41	16:41	17:41	18:41	25.430282, 82.861283
AIRPORT ROAD	1	47	2.40	55	0	7:43	8:43	9:43	10:43	11:43	12:43	13:43	14:43	15:43	16:43	17:43	18:43	25.436178, 82.854178
BABATPUR CHAURAH	2	49	4.80	60	0	7:47	8:47	9:47	10:47	11:47	12:47	13:47	14:47	15:47	16:47	17:47	18:47	25.448526, 82.847216

Babatpur-Kachahari- Cantt-Chandpur-Mohansarai-Mirzamurad

Route NO. (E102)	KM	Total Km	Time	Fare	HLT	First Trip Bus-1 UP65LT 1393	First Trip Bus-2 UP32PN 6687	First Trip Bus-3 UP65KT 5046	First Trip Bus 4 UP32LT 1799	First Trip Bus-5 UP65LT 1398	First Trip Bus 6 UP32PN 8716	Second Trip Bus-1 UP65LT 1393	Second Trip Bus-2 UP32PN 6687	Second Trip Bus-3 UP65KT 5046	Second Trip Bus-4 UP32LT 1799	Second Trip Bus-5 UP65LT 1398	Second Trip Bus-6 UP32PN 8716	Latitude Longitude
BABATPUR CHAURAH	0	0	0.00	0	13	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	25.448526, 82.847216
AIRPORT ROAD	2	2	4.80	10	0	8:04	9:04	10:04	11:04	12:04	13:04	14:04	15:04	16:04	17:04	18:04	19:04	25.436152, 82.854317
SANMUKH HOSPITAL	1	3	2.40	10	0	8:06	9:06	10:06	11:06	12:06	13:06	14:06	15:06	16:06	17:06	18:06	19:06	25.430080, 82.861047
RTO	2	5	4.80	15	0	8:10	9:10	10:10	11:10	12:10	13:10	14:10	15:10	16:10	17:10	18:10	19:10	25.424209, 82.873184
SATOMAHUA	1	6	2.40	15	0	8:12	9:12	10:12	11:12	12:12	13:12	14:12	15:12	16:12	17:12	18:12	19:12	25.411883, 82.885072
KAAZI SARAI	2	8	4.80	20	0	8:16	9:16	10:16	11:16	12:16	13:16	14:16	15:16	16:16	17:16	18:16	19:16	25.398988, 82.892121
HARHUA	1	9	2.40	20	0	8:18	9:18	10:18	11:18	12:18	13:18	14:18	15:18	16:18	17:18	18:18	19:18	25.381367, 82.905753
VYASBAGH	3	12	7.20	25	0	8:25	9:25	10:25	11:25	12:25	13:25	14:25	15:25	16:25	17:25	18:25	19:25	25.372437, 82.921591
B.H.E.L.	1	13	2.40	25	0	8:27	9:27	10:27	11:27	12:27	13:27	14:27	15:27	16:27	17:27	18:27	19:27	25.367641, 82.928782
TARNA	1	14	2.40	25	0	8:29	9:29	10:29	11:29	12:29	13:29	14:29	15:29	16:29	17:29	18:29	19:29	25.362675, 82.939038
AGRASEN MOD	1	15	2.40	30	0	8:31	9:31	10:31	11:31	12:31	13:31	14:31	15:31	16:31	17:31	18:31	19:31	25.359541, 82.961465
GILAT BAZAR	2	17	4.80	30	0	8:35	9:35	10:35	11:35	12:35	13:35	14:35	15:35	16:35	17:35	18:35	19:35	25.354024, 82.965971
BHOJUBEER	1	18	2.40	30	0	8:37	9:37	10:37	11:37	12:37	13:37	14:37	15:37	16:37	17:37	18:37	19:37	25.351406, 82.975131
CIRCUIT HOUSE	1	19	2.40	30	0	8:39	9:39	10:39	11:39	12:39	13:39	14:39	15:39	16:39	17:39	18:39	19:39	25.346475, 82.976838
KACHAHARI	1	20	2.40	35	0	8:41	9:41	10:41	11:41	12:41	13:41	14:41	15:41	16:41	17:41	18:41	19:41	25.342208, 82.981180
NADESAR	1	21	2.40	35	0	8:43	9:43	10:43	11:43	12:43	13:43	14:43	15:43	16:43	17:43	18:43	19:43	25.332815, 82.989755
CANTT RLY STATION	2	23	4.80	35	5	8:52	9:52	10:52	11:52	12:52	13:52	14:52	15:52	16:52	17:52	18:52	19:52	25.324556, 82.986666
LAHARTARA	2	25	4.80	40	0	8:56	9:56	10:56	11:56	12:56	13:56	14:56	15:56	16:56	17:56	18:56	19:56	25.316023, 82.968195
BOULIA	1	26	2.40	40	0	8:58	9:58	10:58	11:58	12:58	13:58	14:58	15:58	16:58	17:58	18:58	19:58	25.315663, 82.964456
CHANDPUR CHAURAH	2	28	4.80	40	0	9:02	10:02	11:02	12:02	13:02	14:02	15:02	16:02	17:02	18:02	19:02	20:02	25.305132, 82.953530
MODHAILA	1	29	2.40	40	0	9:04	10:04	11:04	12:04	13:04	14:04	15:04	16:04	17:04	18:04	19:04	20:04	25.300350, 82.949829
BHULLANPUR	1	30	2.40	40	0	9:06	10:06	11:06	12:06	13:06	14:06	15:06	16:06	17:06	18:06	19:06	20:06	25.296004, 82.942333
ROHANIA	1	31	2.40	45	0	9:08	10:08	11:08	12:08	13:08	14:08	15:08	16:08	17:08	18:08	19:08	20:08	25.289050, 82.927672
JAGATPUR	2	33	4.80	45	0	9:12	10:12	11:12	12:12	13:12	14:12	15:12	16:12	17:12	18:12	19:12	20:12	25.281660, 82.910796
SAHABABAD	2	35	4.80	45	0	9:16	10:16	11:16	12:16	13:16	14:16	15:16	16:16	17:16	18:16	19:16	20:16	25.277763, 82.896476
MOHANSARAI	2	37	4.80	50	0	9:20	10:20	11:20	12:20	13:20	14:20	15:20	16:20	17:20	18:20	19:20	20:20	25.274152, 82.871959
RAJATALAB	3	40	7.20	50	0	9:27	10:27	11:27	12:27	13:27	14:27	15:27	16:27	17:27	18:27	19:27	20:27	25.269067, 82.846402
RAKHAUNA	3	43	7.20	55	0	9:34	10:34	11:34	12:34	13:34	14:34	15:34	16:34	17:34	18:34	19:34	20:34	25.271347, 82.821820
SADHU KI KUTIYA	2	45	4.80	55	0	9:38	10:38	11:38	12:38	13:38	14:38	15:38	16:38	17:38	18:38	19:38	20:38	25.279515, 82.799523
KIT	1	46	2.40	55	0	9:40	10:40	11:40	12:40	13:40	14:40	15:40	16:40	17:40	18:40	19:40	20:40	25.282121, 82.791342
MIRZAMURAD BAZAR	2	48	4.80	55	0	9:44	10:44	11:44	12:44	13:44	14:44	15:44	16:44	17:44	18:44	19:44	20:44	25.285235, 82.776484
MIRZAMURAD ESTATIO	1	49	2.40	60	0	9:46	10:46	11:46	12:46	13:46	14:46	15:46	16:46	17:46	18:46	19:46	20:46	25.284666, 82.780556

Creating a DataFrame with Shapes

- Each table for each route sheet has its own shape from starting stop to last stop.

- The shape is not complete because we don't have linestring of geographical coordinates between two consecutive stops for each table

- Hence, I used MapBox Direction API to generate these coordinates

Shapes.txt features

- Features:
  - shape_id: shape of the route
  - shape_pt_lat: Latitude of a point visited in the trip
  - shape_pt_lon: longitude of a point visited in the trip
  - shape_pt_sequence: Sequence number of the point in the shape

```python
shapes=pd.DataFrame(columns=['shape_id','shape_pt_lat','shape_pt_lon','shape_pt_sequence'])
```

Overall Code

- Iterates over all route sheets and then creates temporary shape data frames for each subtables

- Each temporary shape data frames is appended into main shapes data frame

```python
#Reads through all excel sheets and extract all shapes
shapes=pd.DataFrame(columns=['shape_id','shape_pt_lat','shape_pt_lon','shape_pt_sequence'])
shape_id=0
for route in dfSuperSet.keys():
    #skips the first sheet of excel file
    if route=='50 Bus KM (2)':
        pass
    else:

        c=pd.DataFrame(dfSuperSet[route]).iloc[:-1].dropna(axis=1,how='all')
        c=c.rename(columns={c.columns[0]:'stop_name'})
        #gets the column name for coordinates
        latlon=[col for col in c.columns if 'Latitude' in col][0]


        #Keeps track of null rows between subtables
        index_list=c[c.isnull().all(axis=1)].index
        if route=='Mirzamurad-Chubpur-MARKAND 106':
            index_list=index_list[:3]

        #iterates through each table of the sheet
        for i in range(len(index_list)):
            #checks the first table of the sheet
            if index_list[i]==index_list[0]:
                temp=c.iloc[1:index_list[i]]
                temp['next_stop_coord']=temp[latlon].shift(-1)
                temp['next_stop']=temp['stop_name'].shift(-1)
                temp=temp[temp['next_stop_coord'].notnull()]


                shape_point_sequence=0
                #creates a temporary table for one shape
                temp_shapes=pd.DataFrame(columns=['shape_id','shape_pt_lat','shape_pt_lon','shape_pt_sequence'])
                #iterates through all stops visited in one shape and provides list of points between them
                for k,v in temp.iterrows():
                    #Get coordinates for the origin and destination stops
                    stop_coord = [float(n) for n in v[latlon].split(",")]
                    next_stop_coord=[float(n) for n in v['next_stop_coord'].split(",")]

                    shape_list_of_geometry = shape(stop_coord, next_stop_coord)
                    temp_shapes=temp_shapes.append({'shape_pt_lat':shape_list_of_geometry}, ignore_index=True)

                #create rows for each geographical coordinate visited between stops
                temp_shapes=temp_shapes.explode('shape_pt_lat').reset_index(drop=True)
                temp_shapes[['shape_pt_lon','shape_pt_lat']]=pd.DataFrame(temp_shapes['shape_pt_lat'].tolist(), index= temp_shapes.index)

                temp_shapes['shape_id']=shape_id
                temp_shapes['shape_pt_sequence']=temp_shapes.index
                #appends this table into main shape table
                shapes=shapes.append(temp_shapes).reset_index(drop=True)
                shape_id+=1



            #skips consecutive empty lines
            elif c.iloc[index_list[i-1]+1:index_list[i]].empty:
                pass
            else:
                #checks the later tables of the sheet
                temp=c.iloc[index_list[i-1]+4:index_list[i]].reset_index(drop=True)
                #works same as first table
                temp['next_stop_coord']=temp[latlon].shift(-1)
                temp['next_stop']=temp['stop_name'].shift(-1)
                temp=temp[temp['next_stop_coord'].notnull()]


                shape_point_sequence=0
                temp_shapes=pd.DataFrame(columns=['shape_id','shape_pt_lat','shape_pt_lon','shape_pt_sequence'])
                for k,v in temp.iterrows():
                    stop_coord = [float(n) for n in v[latlon].split(",")]
                    next_stop_coord=[float(n) for n in v['next_stop_coord'].split(",")]

                    shape_list_of_geometry = shape(stop_coord, next_stop_coord)
                    temp_shapes=temp_shapes.append({'shape_pt_lat':shape_list_of_geometry}, ignore_index=True)

                temp_shapes=temp_shapes.explode('shape_pt_lat').reset_index(drop=True)
                temp_shapes[['shape_pt_lon','shape_pt_lat']]=pd.DataFrame(temp_shapes['shape_pt_lat'].tolist(), index= temp_shapes.index)
                temp_shapes['shape_id']=shape_id
                temp_shapes['shape_pt_sequence']=temp_shapes.index
                shapes=shapes.append(temp_shapes).reset_index(drop=True)
                shape_id+=1
```

Code Snippets for creating a Data Frame with Shapes

- This code snippet converts an excel sheet containing information of one route into a pandas data frame. index_list stores indices where the rows are null (separating two subtables)

- DfSuperSet stores all excel sheets of all routes

```python
c=pd.DataFrame(dfSuperSet[route]).iloc[:-1].dropna(axis=1,how='all')
c=c.rename(columns={c.columns[0]:'stop_name'})
#gets the column name for coordinates
latlon=[col for col in c.columns if 'Latitude' in col][0]


#Keeps track of null rows between subtables
index_list=c[c.isnull().all(axis=1)].index
if route=='Mirzamurad-Chubpur-MARKAND 106':
    index_list=index_list[:3]
```

Code Snippets for creating a Data Frame with Shapes

- This function takes the coordinates of two consecutive stops of a shape and generates a list of geographical coordinates it will visit in between the two stops

```python
#Generates a list of geographical coordinates representing linestring from origin stop to destination stop
def shape(stop_coord,next_stop_coord):
    x = requests.get("https://api.mapbox.com/directions/v5/mapbox/driving/"+str(stop_coord[1])+"%2C"+str(stop_coord[0])+\
                     "%3B"+str(next_stop_coord[1])+"%2C"+str(next_stop_coord[0])+\
                     "?alternatives=true&geometries=geojson&language=en&overview=simplified&steps=true&access_token=pk.eyJ1IjoienVscWFyYSIsImEiOiJjbGVkNWxmMDMwMzN1M3lzMmF6bGkzZm9z
    return x.json()['routes'][0]['geometry']['coordinates']
```

Code Snippets for creating a Data Frame with Shapes

- This code snippet iterates through each table and call shape function for each consecutive stops. It then stores all the geographical coordinates between stops as rows and append it to main shapes data frame

```python
#creates a temporary table for one shape
temp_shapes=pd.DataFrame(columns=['shape_id','shape_pt_lat','shape_pt_lon','shape_pt_sequence'])
#iterates through all stops visited in one shape and provides list of points between them
for k,v in temp.iterrows():
    #Get coordinates for the origin and destination stops
        stop_coord = [float(n) for n in v[latlon].split(",")]
        next_stop_coord=[float(n) for n in v['next_stop_coord'].split(",")]

        shape_list_of_geometry = shape(stop_coord, next_stop_coord)
        temp_shapes=temp_shapes.append({'shape_pt_lat':shape_list_of_geometry}, ignore_index=True)

#create rows for each geographical coordinate visited between stops
temp_shapes=temp_shapes.explode('shape_pt_lat').reset_index(drop=True)
temp_shapes[['shape_pt_lon','shape_pt_lat']]=pd.DataFrame(temp_shapes['shape_pt_lat'].tolist(), index= temp_shapes.index)

temp_shapes['shape_id']=shape_id
temp_shapes['shape_pt_sequence']=temp_shapes.index
#appends this table into main shape table
shapes=shapes.append(temp_shapes).reset_index(drop=True)
shape_id+=1
```

Shapes Data Frame

	shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence
0	0	25.284410	82.780492	0
1	0	25.282874	82.788947	1
2	0	25.281990	82.792961	2
3	0	25.283655	82.793669	3
4	0	25.283546	82.794604	4
...	...	...	...	...
8272	32	25.285850	82.776031	529
8273	32	25.285274	82.775819	530
8274	32	25.285149	82.776464	531
8275	32	25.285149	82.776464	532
8276	32	25.284410	82.780492	533

8277 rows × 4 columns