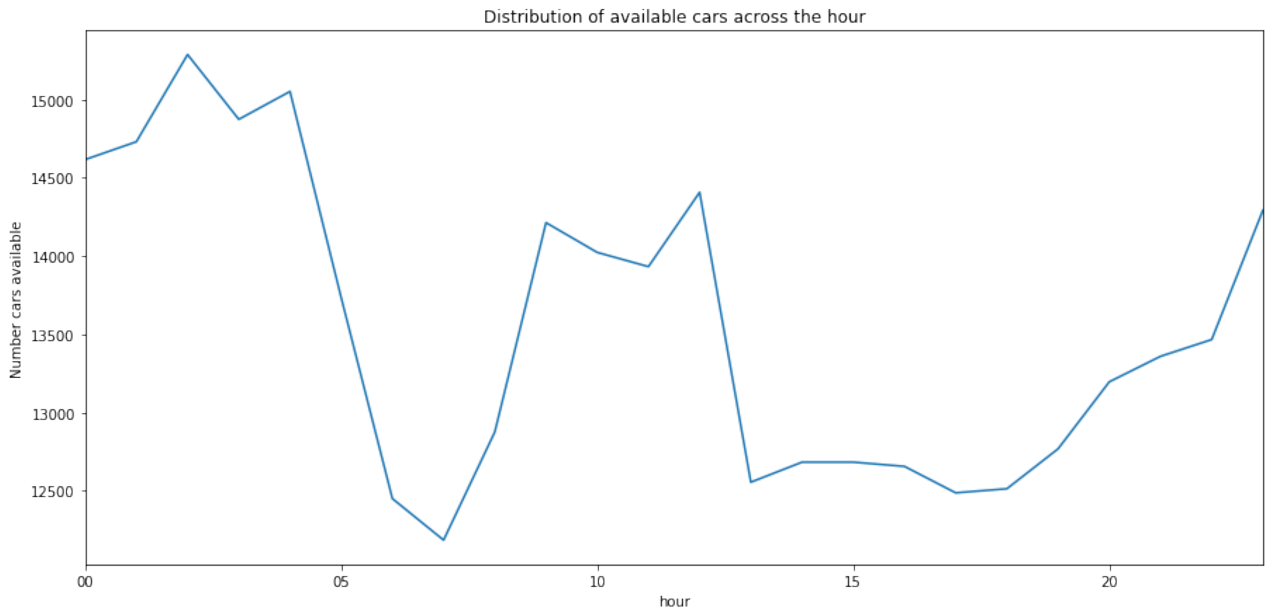
We can see from distribution above Fridays (11th january and 18th january) seems to be days where available cars are minimum and hence peak of demand comes on this day. Also we can see somewhat weekly pattern in demand here.

In [152]:
```
fig, ax = plt.subplots(figsize=(15,7))
df1.groupby(['hour']).sum()['total_cars'].plot(ax=ax)
ax.set_xlabel('hour')
ax.set_ylabel('Number cars available')
ax.set_title('Distribution of available cars across the hour')
```

Out[152]:
```
Text(0.5, 1.0, 'Distribution of available cars across the hour')
```



Learning from image 3

We can see from graph above, number of cars available dips in morning and evening commute hours. In morning, 6 AM to 8 AM and in evening 3 PM to 7 PM seems to be peak hours in terms of demand of these cars.

In [153]:
```
df1_zero = df1[df1['total_cars'] == 0]
df1_zero_grouped = df1_zero.groupby(['geohash']).count()['total_cars']

df1_non_zero = df1[df1['total_cars'] > 0]
df1_non_zero_grouped = df1_non_zero.groupby(['geohash']).count()['total_
```

In [154]:
```
df1_zero_grouped.head(5)
```

Out[154]:
```
geohash
sv8wnz     304
sv8wqb    1579
sv8wqc    1566
sv8wqf    5867
sv8wqg    4785
Name: total_cars, dtype: int64
```

In [155]:
```
df1_non_zero_grouped.head(5)
```

Out[155]:
```
geohash
sv8tz0       3
sv8wng      14
sv8wnw     731
sv8wnx    2146
sv8wny       1
Name: total_cars, dtype: int64
```

In [156]:
```
df1_geohash_merged = df1_zero_grouped.to_frame().merge(df1_non_zero_grou
```

In [157]:
```
#df1_geohash_merged.head(5)
df1_geohash_merged.rename(columns={"total_cars_x": "instances with 0 car
df1_geohash_merged.head(5)
```

Out[157]:

	total_cars_x	total_cars_y
geohash		
sv8wnz	304	2260
sv8wqb	1579	4340
sv8wqc	1566	4848
sv8wqf	5867	6759
sv8wqg	4785	5121

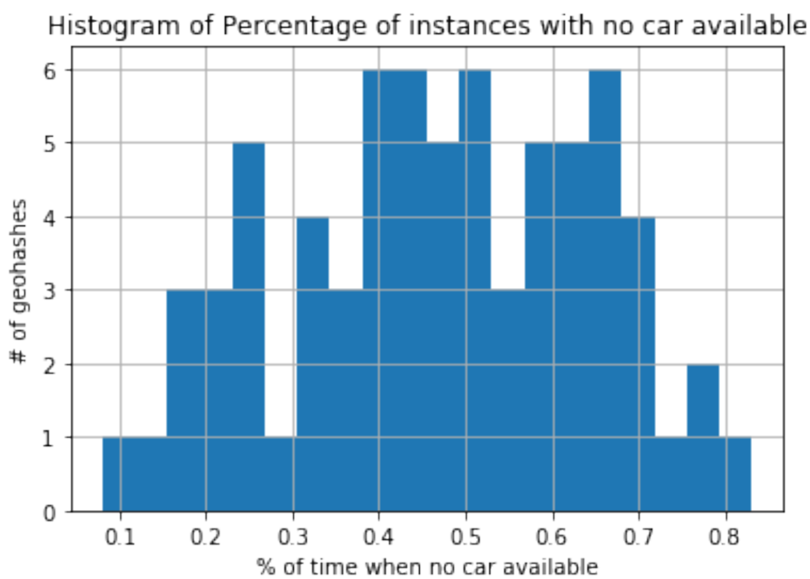
In [158]:
```
df1_geohash_merged['zero cars %'] = df1_geohash_merged['total_cars_x'] /
df1_geohash_merged.head(5)
```

Out[158]:

	total_cars_x	total_cars_y	zero cars %
geohash			
sv8wnz	304	2260	0.118565
sv8wqb	1579	4340	0.266768
sv8wqc	1566	4848	0.244153
sv8wqf	5867	6759	0.464676
sv8wqg	4785	5121	0.483041

In [159]:
```
import pylab as pl
df1_geohash_merged['zero cars %'].hist(bins = 20)
pl.title("Histogram of Percentage of instances with no car available")
pl.xlabel("% of time when no car available")
pl.ylabel("# of geohashes")
```

Out[159]: Text(0, 0.5, '# of geohashes')



Learning from image 4

We can see there are few geohashes which have more than 70% of times (right of the histogram above) no car being available. These are the locations where either demand is typically high and/or theses locations do not see people parking cars. There seems to be some optimizaiton scope if we can move some cars to these locations.

In [ ]: