**Leipziger str south:**

**Before removing:**

**Variance 145.705501**

**Expon sse: 22771.40234860699, rejected null hypothesis with 95% C.I**

**Lognorm sse: 10547.130880170222, rejected null hypothesis with 95% C.I**

**Norm sse : 43318.018614922636, rejected null hypothesis with 95% C.I**

**Data < 15**

exp: close fit in qq plot but Rejected null hypothesis with 95% confidence interval as chi value is not less than tabulated(close to theoretical ), sse = 96.99359997385118

Lognormal: Failed to reject null hypothesis with 95% confidence interval at k = 34 chi =40.55014279825781 < 44.985 (df = 31) (not only 34but for most of them), sse= 294.0773265479426

Norm:Rejected null hypothesis with 95% confidence interval, chi value is not less than tabulated, sse =959.3228009715183

Variance = 7.719009

**Data < 25**

exp:Rejected null hypothesis with 95% confidence interval but close fit in qq plot, sse = 908.1043538692305

Lognormal: failed to reject null hypothesis at k = 80, chi = 96.8853491051358< 98.484 (df = 77)and for few other values after 81, close fit in qq plot, sse = 1139.8242136414551

Norm: No proper fit and Rejected null hypothesis with 95% confidence interval, sse = 3620.8693617585723

Variance = 19.972214

**Data< 30**

exp: close fit but Rejected null hypothesis with 95% confidence interval , chi value is not less than tabulated for any k , sse = 1506.8272257584504

Lognormal:failed to reject null hypothesis with 95% confidence level at k=83 , chi = 97.01192955789713< 101.88(df = 80), sse =1458.0231606620443

Norm: Rejected null hypothesis with 95% confidence interval, chi value is not less than tabulated for any k , sse = 5141.926255269538

Variance = 26.079071

**Data < 40**

exp: Df< 30 exp:Rejected null hypothesis with 95% confidence interval, chi value is not less than tabulated for any k , sse = 4586.7407320703205

Lognormal: failed to reject null hypothesis at(k=91) = chi = 108.835 < 110.90(df = 88), sse = 3141.7580461553866

Norm:Rejected null hypothesis with 95% confidence interval, chi value is not less than tabulated for any k , sse = 11457.69245570545

variance = 48.390222

CONCLUSION : lognormal Data<15 would be a best choice as it failed to reject null hypothesis and qq plot fits well with slope =1 for considerable range.

As the variance decreased to a greater extent the mean is shifted and does not consider the data which has very less impact on the system.

**Leipziger str north:**

**Before cleaning**

**Variance 123.529623**

**Expon sse : 7926.433300465245, rejected hypothesis**

**log norm sse: 5956.732931408156, failed to reject hypothesis (ex : at k = 25 chi calculated< critical, for most of the values after k = 25 also failed to reject null hypothesis)**

**Norm sse : 30251.692089104705, rejected hypothesis**

**Data < 15**

exp: Rejected null hypothesis but a good fit with slope almost 1 for quite a good range

Lognormal: Rejected null hypothesis but a good fit with slope almost 1 for quite a good range

Norm: qq plot is a close fit but rejected null hypothesis for every k.

Variance = 11.513085

**Data < 25**

exp: very close fit and failed to reject null hypothesis at k = 99 chi = 117.5105698172637<119.871 and for few other values of k as well, sse = 575.0.

Lognormal: (at k = 69) chi = 82.65121610957159< 85.97 and also few other k values after

Fail to reject null hypothesis, sse= 4102

Norm: Rejected null hypothesis and also not a good fit as sse is very high compared to above 2.

**Data< 30**

exp: very good fit and failed to reject null hypothesis after k = 109 chi = 129.86242772202303<135 (df = 106)and few other values of k > 109, sse = 740, correlation is strong enough.

Lognormal: good fit and fails to reject null hypothesis at k = 77 chi = 88.88832567719258< 95.08 and also for few values of k after 77., sse = 6351

Normal : Not a good fit and rejected hypothesis.

Variance = 40.026998

**Data < 40**

exp: good fit and chi value is not less than tabulated, sse =1527

Lognormal: not an appropriate fit but failed to reject null hypothesis at(k=62) = chi = 74.80234693497117< 77.931 (df = 59) and few other values of k > 61 as well, sse = 10384.463667909342 but the correlation is not strong enough.

Norm: chi value is not less than tabulated

Variance = 61.776509

CONCLUSION:

1. The distribution would be log normal for raw data as from the chi square test it is clear that the null hypothesis is not rejected with 95% C.I 2 and also the histogram seems to be lognormal.
2. But In order to optimize the squared errors data cleaning is done and the distribution will be exponential for data<30 or data < 25 as the sum of squared error is also considerably less compared to other ranges.

.Suggestion: Go with full data and with log normal distribution.