EUROPEAN INSTITUTE OF INNOVATION, ENTREPRENEURSHIP, AND TECHNOLOGY

(EIIET)

Berlin Campus

Fundamentals of Statistics

Descriptive Data Analytics and Comparative Market Performance

Practical Analytical Report (Data Analysis Project) – Assignment Option 1

Companion spreadsheet: 5. Statistics Option 1 Spreadsheet

November 2026

Student name: Muhammad Sameer Randhawa

Student ID: DE240359

Table of Contents

Executive Summary

1. Introduction

2. Theoretical Framework & Statistical Laws

3. Data Analysis & Statistical Calculations

4. Discussion & Economic Implications

5. Conclusion & Recommendations

References

Appendices (Spreadsheet Data & Computational Output)

Executive Summary

This report uses descriptive statistics to compare the weekly sales of two branches of a fictional coffee chain over 52 weeks. The data are illustrative and were analysed in a spreadsheet using standard Excel functions (Appendix). **Branch A (city centre)** sold more on average (€12,285 per week) and was very stable: its standard deviation was €1,082, a coefficient of variation of 8.8%. Its distribution is right-skewed (skewness 1.01) because three event weeks produced unusually high sales. **Branch B (station)** averaged €11,164 with almost twice the dispersion (standard deviation €2,079, CV 18.6%) and a mild left skew (-0.40) caused by weak winter weeks and one severe disruption. Branch B fell below an assumed cost-covering level of €9,500 in 15 of 52 weeks, all of them in winter, while Branch A never did. The executive recommendations are to plan staffing and stock for Branch B on a seasonal basis, to build an event calendar for Branch A that repeats its best weeks, and to monitor both branches monthly with the median, IQR and CV rather than the mean alone.

1. Introduction

Managers often compare business units using a single figure such as average sales. Averages, however, hide how much results vary from week to week, and variability matters for staffing, stock planning, cash flow and risk. Descriptive statistics summarise a dataset by its centre, spread and shape, giving decision-makers a fuller and more reliable picture (Anderson et al., 2020).

The objective of this report is to evaluate the operational performance of two branches using descriptive methods and to turn the results into evidence-based recommendations. The dataset contains weekly sales in euros for Branch A, located in a city centre, and Branch B, located in a railway station, for weeks 1–52 of one year. The data are fictional and were created for teaching purposes. Section 2 sets out the statistical concepts, Section 3 presents the analysis, Section 4 discusses economic implications and Section 5 gives conclusions and recommendations.

2. Theoretical Framework & Statistical Laws

2.1 Frequency distributions and graphs

A frequency distribution groups observations into classes and counts how many fall into each class. It reveals the shape of the data: whether values cluster in the middle, spread widely or have long tails. Histograms display frequency distributions, line charts show values over time

and box plots summarise the median, quartiles and extreme values in one picture (Tukey, 1977).

2.2 Measures of central tendency

The **mean** is the arithmetic average, $\bar{x} = \Sigma x \div n$. It uses every observation but is pulled towards extreme values. The **median** is the middle value of the ordered data and is resistant to outliers. The **mode** is the most frequent value; for continuous data such as sales it is more useful to report the **modal class** of a frequency distribution. When mean > median, the distribution is usually right-skewed; when mean < median, it is usually left-skewed.

2.3 Measures of dispersion and shape

The **range** is the maximum minus the minimum. The sample **variance**, $s^2 = \Sigma(x - \bar{x})^2 \div (n - 1)$, measures the average squared deviation from the mean, and the **standard deviation** s is its square root, expressed in the original units. The **coefficient of variation**, $CV = s \div \bar{x}$, expresses dispersion relative to the mean and allows units of different size to be compared. The **interquartile range**, $IQR = Q3 - Q1$, covers the middle 50% of observations and is not affected by outliers. **Skewness** measures asymmetry: positive values indicate a longer right tail and negative values a longer left tail (Joanes & Gill, 1998). By Chebyshev's theorem, at least 75% of any dataset lies within two standard deviations of the mean, so the standard deviation gives a usable risk band even when data are not normally distributed.

2.4 Descriptive analytics in business practice

Descriptive analytics is the first layer of business analytics: it answers the question "what happened?" before predictive or prescriptive methods ask what will happen or what should be done (Levine et al., 2021). Retail chains compare stores by average basket size and its dispersion, hotels track the spread of daily occupancy to set staffing levels, and service businesses use the distribution of waiting or delivery times rather than the average to set service promises. In each case, measures of spread and shape reveal risks and opportunities that the mean conceals. Because these calculations are available in standard spreadsheet software, even small firms can apply them to their own transaction data.

3. Data Analysis & Statistical Calculations

3.1 Data over time

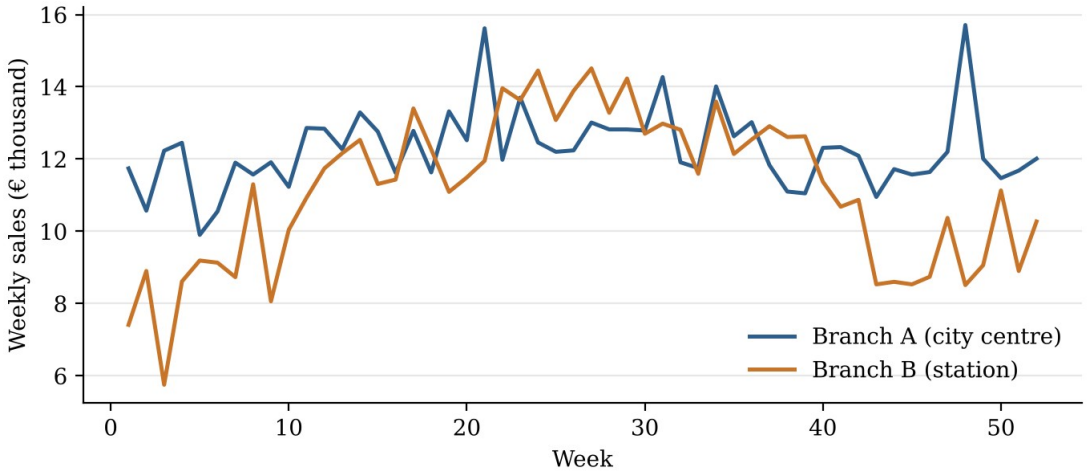


Figure 1. Weekly sales by branch, weeks 1–52 (fictional data)

Figure 1 shows two different patterns. Branch A moves in a narrow band around €12,000 with three clear spikes in weeks 21, 34 and 48 (between €14,010 and €15,710), which coincide with city events. Branch B follows a strong seasonal wave: sales average €13,298 in the summer weeks 23–35 but only €8,712 in weeks 1–10. Week 3 at Branch B (€5,750) reflects a rail strike.

3.2 Frequency distribution

Table 1. Frequency distribution of weekly sales

Sales class	Branch A (weeks)	Branch A (%)	Branch B (weeks)	Branch B (%)
€4,000 – under €6,000	0	0.0%	1	1.9%
€6,000 – under €8,000	0	0.0%	1	1.9%
€8,000 – under €10,000	1	1.9%	13	25.0%
€10,000 – under €12,000	21	40.4%	16	30.8%
€12,000 – under €14,000	26	50.0%	18	34.6%
€14,000 – under €16,000	4	7.7%	3	5.8%
€16,000 – under €18,000	0	0.0%	0	0.0%
Total	52	100.0%	52	100.0%

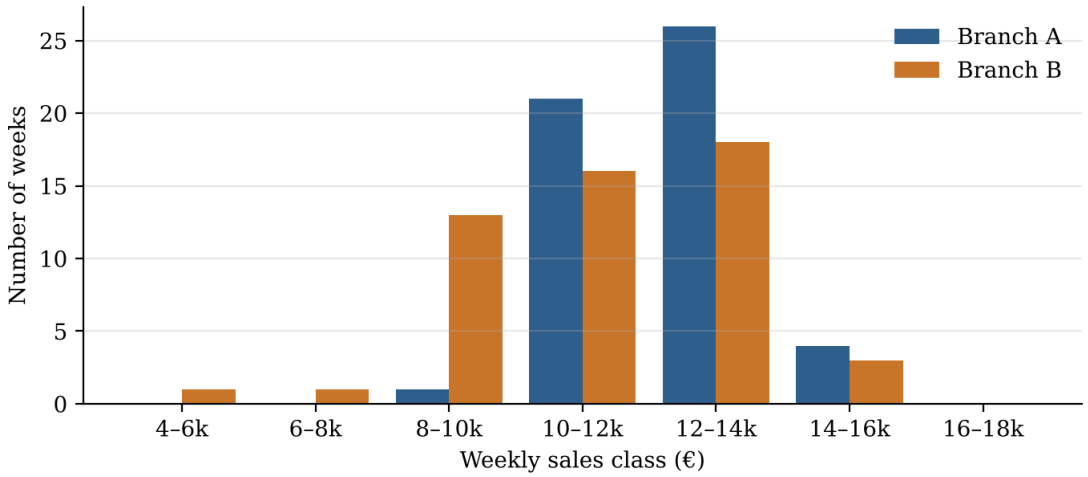


Figure 2. Histogram of weekly sales

Branch A is concentrated: 47 of 52 weeks (90.4%) fall between €10,000 and €14,000. Branch B is spread across five classes, with 15 weeks (28.8%) below €10,000. For both branches the **modal class** is €12,000 – under €14,000 (26 weeks for A, 18 for B). Because sales values were recorded to the nearest €10, individual values rarely repeat, so the modal class is the meaningful measure of the mode here.

3.3 Summary statistics

Table 2. Descriptive statistics (computed in Excel; see Appendix)

Measure	Branch A (city centre)	Branch B (station)
Mean	€12,285	€11,164
Median	€12,200	€11,400
Modal class	€12,000–14,000	€12,000–14,000
Minimum	€9,900	€5,750
Maximum	€15,710	€14,510
Range	€5,810	€8,760
Variance (s²)	1,171,272	4,321,031
Standard deviation (s)	€1,082	€2,079
Coefficient of variation	8.8%	18.6%
Q1	€11,670	€9,112
Q3	€12,798	€12,728
Interquartile range	€1,128	€3,615
Skewness	1.01	-0.40
Annual total	€638,840	€580,510

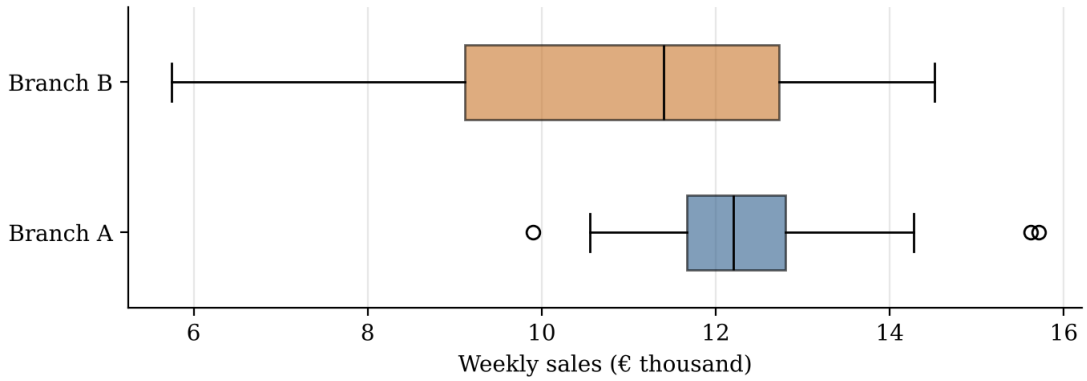


Figure 3. Box plots of weekly sales

3.4 Interpretation

Centre. Branch A's mean (€12,285) is €1,122 higher than Branch B's. For Branch A the mean lies above the median (€12,200), consistent with right skew from the event weeks; without them its mean would be about €12,112. For Branch B the mean lies below the median (€11,400), consistent with a left tail of weak winter weeks.

Spread. Branch B's standard deviation (€2,079) is almost twice Branch A's (€1,082), and its CV of 18.6% against 8.8% confirms that the difference is not just due to scale. The IQR shows the gap even more clearly: the middle half of Branch B's weeks span €3,615, more than three times Branch A's €1,128. Applying Chebyshev's theorem, at least 75% of Branch B's weeks should fall within about €7,006–€15,321, a band too wide for fixed staffing.

Shape. Branch A's skewness of 1.01 is moderately positive: upside surprises are larger than downside ones. Branch B's -0.40 is mildly negative: its risk lies in sudden drops. In the box plot, Branch A's event weeks appear as high outliers, while Branch B's strike week appears as a low outlier.

4. Discussion & Economic Implications

The analysis shows how descriptive statistics answer three common business questions.

- **Sales performance.** Branch A earned €58,330 more over the year. A comparison of means alone would suggest only a modest gap, but the distribution shows that Branch A is also far more predictable.
- **Cost efficiency.** Staff rosters and stock orders are usually planned around a typical week. At Branch A, planning to the median works well because 50% of weeks lie within a band of about €1,100. At Branch B, a fixed plan would cause overstaffing and waste in winter

and lost sales in summer. Assuming the branch needs about €9,500 a week to cover its costs, Branch B fell short in 15 weeks, all between weeks 1–9 and 43–51, which points to a seasonal cost problem rather than a general one.

- **Customer trends.** Branch A's positive skew indicates untapped demand on event days: the three best weeks were roughly €2,000–3,500 above normal. Branch B's seasonality follows commuter and tourist traffic at the station, which the business cannot control but can forecast.

In risk terms, the standard deviation and CV act as measures of operational volatility. Higher volatility raises the cash buffer a branch needs and makes its profit more sensitive to fixed costs, which matters for decisions on leases, staffing contracts and expansion. A business that only reported average sales would miss these points.

The analysis has limitations. The data are fictional and cover one year, so seasonal patterns cannot be confirmed across years. Descriptive statistics describe but do not explain causes; regression or time-series methods would be needed to quantify the effect of events or weather. Finally, sales are not profit: margins may differ between the branches.

5. Conclusion & Recommendations

Branch A delivers higher, stable sales with occasional positive spikes; Branch B delivers lower, volatile and strongly seasonal sales with occasional sharp drops. Using measures of centre, spread and shape together gives management a much more accurate basis for planning than averages alone. For the executive decision-maker, the following actions are recommended:

1. **Seasonal planning for Branch B:** use separate summer and winter staffing and stock plans, and consider reduced opening hours in weeks 1–9 and 43–52.
2. **Event strategy for Branch A:** build a calendar of city events and plan extra staff and promotions for them to repeat the high-sales weeks.
3. **Risk buffer:** hold a larger cash and stock buffer for Branch B, sized at about one standard deviation of weekly sales (≈ €2,100).
4. **Better reporting:** add the median, IQR and coefficient of variation to the monthly management dashboard alongside the mean.
5. **Further analysis:** collect two or more years of data and add transaction counts and margins to separate customer numbers from spend per customer.

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2020). *Statistics for business and economics* (14th ed.). Cengage.

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 183–189.

Levine, D. M., Szabat, K. A., & Stephan, D. F. (2021). *Statistics for managers using Microsoft Excel* (9th ed.). Pearson.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Appendices (Spreadsheet Data & Computational Output)

Appendix A – Raw data: weekly sales (€, fictional)

Week	Branch A	Branch B	Week	Branch A	Branch B
1	11,740	7,410	27	13,010	14,510
2	10,570	8,900	28	12,820	13,280
3	12,230	5,750	29	12,820	14,230
4	12,450	8,610	30	12,790	12,700
5	9,900	9,190	31	14,270	12,980
6	10,550	9,130	32	11,910	12,810
7	11,900	8,730	33	11,750	11,590
8	11,570	11,300	34	14,010	13,590
9	11,910	8,060	35	12,630	12,140
10	11,230	10,040	36	13,020	12,540
11	12,860	10,930	37	11,830	12,910
12	12,840	11,740	38	11,100	12,610
13	12,270	12,160	39	11,050	12,630
14	13,290	12,530	40	12,310	11,370
15	12,760	11,310	41	12,330	10,680
16	11,620	11,430	42	12,090	10,870
17	12,780	13,400	43	10,950	8,530
18	11,630	12,270	44	11,720	8,600
19	13,320	11,090	45	11,570	8,530
20	12,520	11,490	46	11,640	8,740
21	15,620	11,950	47	12,200	10,370
22	11,980	13,960	48	15,710	8,510
23	13,700	13,630	49	12,010	9,060
24	12,460	14,450	50	11,470	11,130
25	12,200	13,080	51	11,680	8,900
26	12,240	13,890	52	12,010	10,270

Appendix B – Spreadsheet formulas used

Measure	Excel formula (Branch A, data in B4:B55)
Mean	=AVERAGE(B4:B55)
Median	=MEDIAN(B4:B55)
Sample variance	=VAR.S(B4:B55)
Standard deviation	=STDEV.S(B4:B55)
Coefficient of variation	=STDEV.S(B4:B55)/AVERAGE(B4:B55)
Quartiles	=QUARTILE(B4:B55,1) and =QUARTILE(B4:B55,3)
Skewness	=SKEW(B4:B55)
Class frequency	=COUNTIFS(B4:B55,">=12000",B4:B55,"<14000")

The companion workbook Statistics*Option1*Spreadsheet.xlsx contains the data and all formulas; changing the data recalculates every statistic.