Final Report : Stepping into New Territory: Adidas US Sales Analysis and the North Korea Opportunity

GROUP 7 : Aditya Kumar Thota, Andre Andrea, Damilola Ayodeji, Sandeep Khanal

1. Background

In the pursuit of expanding its global presence, Adidas is considering the challenging prospect of entering the North Korean market, a region marked by profound isolation and distinctive economic and political circumstances. Leveraging its reputation for data-driven decision-making, Adidas is embarking on an analytical journey. The project aims to thoroughly analyze historical sales data, capturing the nuances of the present US market, while aligning with the broader strategic objectives of the company. The ultimate goal is to provide actionable insights and a comprehensive assessment of the feasibility of establishing a profitable business presence in North Korea.

1.2 Questions

1. Which product line has the greatest number of sales?
2. Which retailer has the most market presence based on the sales method?
3. Is there any significant difference in total sales based on the sales method?
4. Is there any correlation between two numerical variables which could help us understand the relationship between them?
5. If we were to setup the business, will it be profitable, if not what reasons are to be considered before the setup?

1.3 Project Significance

The project holds significance on multiple fronts. To begin, it underscores Adidas's dedication to informed decision-making, reflecting its adaptability in considering market entry into North Korea, a region defined by intricate isolation and unique economic and political circumstances. If successful, this initiative has the potential to diversify Adidas's global market presence and amplify its brand reputation. Additionally, it unlocks vital insights into the North Korean market, a notably data-scarce environment, with applications not only for Adidas but also for businesses and governments seeking economic intelligence about the region.

2. Demographic Comparison between USA and North Korea[1]

Indicators	United States (2021)	North Korea (2021)
Age structure	**0-14 years: 18.46%** (male 31,374,555/female 30,034,371) **15-24 years: 12.91%** (male 21,931,368/female 21,006,463) **25-54 years: 38.92%** (male 64,893,670/female 64,564,565) **55-64 years: 12.86%** (male 20,690,736/female 22,091,808) **65 years and over: 16.85%** (male 25,014,147/female 31,037,419)	**0-14 years: 20.33%** (male 2,680,145/female 2,571,334) **15-24 years: 14.39%** (male 1,873,814/female 1,842,269) **25-54 years: 43.77%** (male 5,671,900/female 5,633,861) **55-64 years: 11.77%** (male 1,454,000/female 1,585,830) **65 years and over: 9.75%** (male 878,176/female 1,640,031) (2021 est.)
Obesity - adult prevalence rate	**41.9% (2023)**	**8.3% (2023)**

1. **Target Audience:** Understanding the age structure in North Korea, with a significant portion of the population falling into the 0-14 years and 15-24 years age groups, can be crucial for Adidas. It suggests a potential market of young consumers who may be interested in sports and athletic products.
2. **Health and Wellness:** The relatively low obesity prevalence rate (8.3%) in North Korea implies that the population may generally have a healthier lifestyle. This could align with Adidas's products, promoting fitness and wellness.
3. **Market Opportunity:** A lower prevalence of obesity may indicate an interest in physical fitness and sports-related activities, potentially creating a market opportunity for sports apparel and equipment.

3.Data Analysis[2]

The Adidas sales dataset incorporates information regarding sales of Adidas products. This dataset includes details such as the total sales revenue, the number of units sold, the product categories, the location of the sales, and any other relevant information. The dataset contains 9649 rows and 13 columns. The main attributes of the Adidas US sales dataset include Retailer, Retailer ID, Invoice Date, Region, State, City, Product, Price per Unit, Units Sold, Total Sales, Operating Profit, Operating Margin, and Sales Method. The dataset contains float data types and has no missing or null values. This dataset serves as a foundation for detailed analysis and decision-making regarding sales, profitability, and product performance in New York City over the two-month period.

[1] https://www.indexmundi.com/factbook/compare/united-states.north-korea
[2] https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset d

3.1 Understanding the dataset

The provided data offers a range of descriptive statistics for various variables, including "Price per Unit," "Units Sold," "Total Sales," "Operating Profit," and "Operating Margin." These insights can be generated from the data:

1. Central Tendencies:

    i.   The mean (average) "Price per Unit" is approximately $45.22, providing an idea of the typical price point.
    ii. The mean "Units Sold" is around 256.93, indicating the average number of units sold.
    iii. The mean "Total Sales" is approximately $93,273.44, representing the average total sales revenue.
    iv. The mean "Operating Profit" is about $34,425.24, signifying the average profit.
    v.   The mean "Operating Margin" is 0.42, which is the average profit margin as a percentage.

2. Variability:

    i.   The standard error measures the standard deviation of the sample means, indicating how much the sample means might vary from the population mean.
    ii. The standard deviation quantifies the degree of variability in the dataset. For instance, the "Price per Unit" has a relatively high standard deviation of approximately $14.70, suggesting significant price variability.
   iii. Sample variance offers insights into the spread of the data points.

3. Distribution:

    i.   The median is a measure of central tendency and indicates that half of the data points fall below $45.00 for "Price per Unit."
    ii. The mode represents the most frequently occurring values in the dataset.

These insights are valuable for understanding the central tendencies, variability, and distribution of the data, providing a basis for further analysis, decision-making, and comparisons within the dataset. (see below for more information pertaining to Descriptive Summary)

Descriptive summary of the Numerical variables

Descriptive Values	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin
Mean	$45.22	256.93	$93,273.44	$34,425.24	$0.42
Standard Error	$0.15	2.18	$1,444.74	$551.70	$0.00
Median	$45.00	176.00	$9,576.00	$4,371.42	$0.41
Mode	50	225.00	100000	21000	0.35
Standard deviation	14.70463501	214.24	141908.6619	54190.30513	0.097192381
Sample Variance	216.2487046	45903.93	20140155804	2936893574	0.009447338
Minimum	$7.00	0.00	$0.00	$0.00	$0.10
Maximum	$110.00	1275.00	$825,000.00	$390,000.00	$0.80
Sum	$436,250.00	2478861.00	$899,902,125.00	$332,134,761.45	$4,081.02
Count	9648	9648.00	9648	9648	9648

3.2 Preliminary data analysis
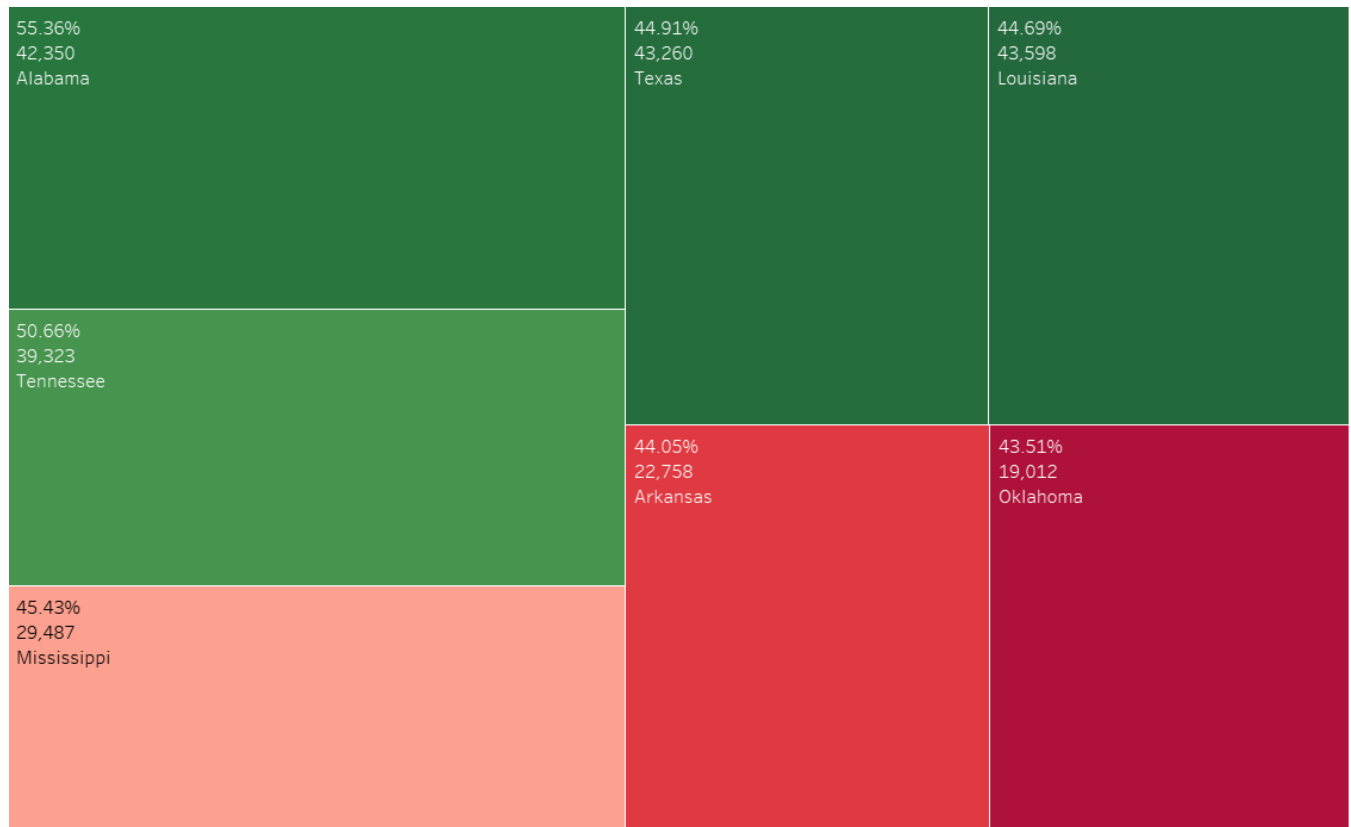
Here we look at the state wise average profit margins for top 7 states, here we can derive that Alabama has the most profit margin compared to other states, next comes Tennessee with 50.66% Profit margin.
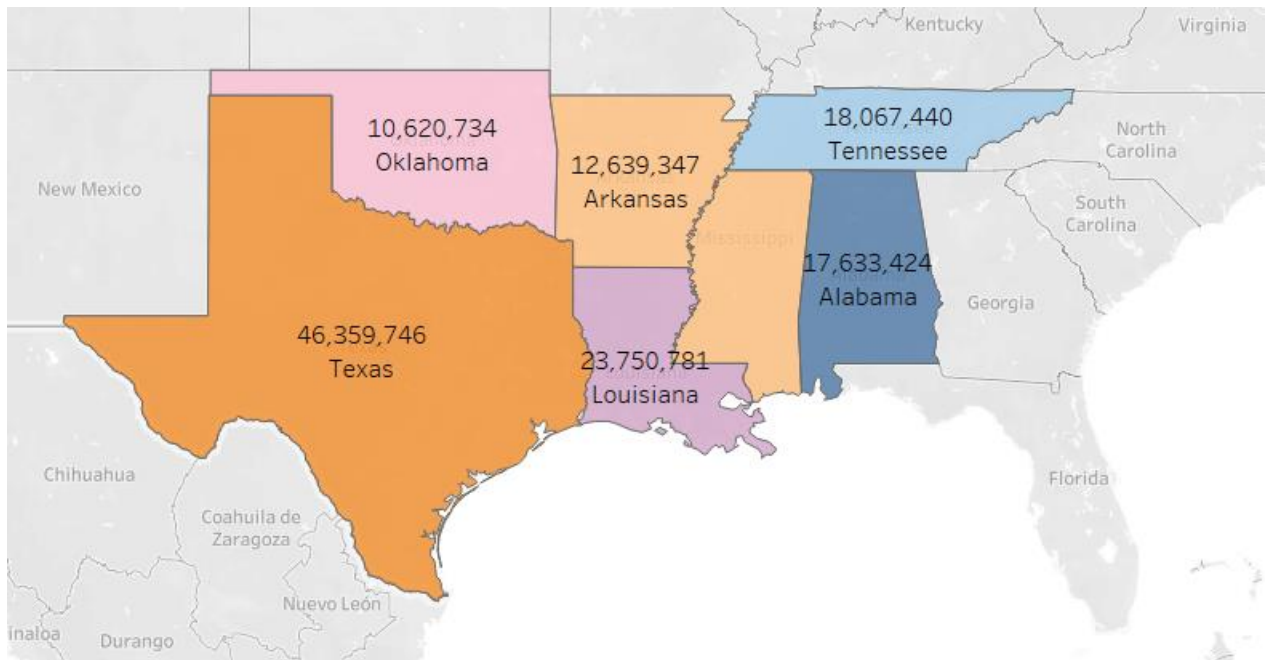
The map shows the total sales distribution across all the states, Texas leading with 46,359,746 in sales, followed by Louisiana with 23,750,781. This distribution shows which state has the majority sales.

4.Statistical Methods

4.1 Correlation Analysis

Variables	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin
Price per Unit	-	0.27	0.44	0.39	-0.14
Units Sold	0.27	-	**0.91**	**0.89**	-0.31
Total Sales	0.44	**0.91**	-	**0.96**	-0.36
Operating Profit	0.39	**0.89**	**0.96**	-	-0.21
Operating Margin	-0.14	-0.31	-0.36	-0.21	-

The correlation analysis reveals important relationships between variables:

1. **"Price per Unit"** and **"Units Sold"** have a weak positive correlation **(0.27)**, meaning that higher prices per unit are associated with more units sold.

2. **"Price per Unit"** and **"Total Sales"** show a moderate positive correlation **(0.44)**, indicating that higher unit prices tend to result in higher total sales.

3. **"Units Sold"** and **"Total Sales"** exhibit a strong positive correlation **(0.91)**, suggesting that as units sold increase, total sales significantly rise.

4. **"Units Sold"** and **"Operating Profit"** are highly positively correlated **(0.89)**, meaning higher unit sales are associated with significantly higher operating profits.

5. **"Total Sales"** and **"Operating Profit"** are very strongly correlated **(0.96)**, implying that higher sales result in significantly higher operating profits.

6. **"Operating Profit"** and **"Operating Margin"** have a negative correlation **(-0.21)**, indicating that as operating profit increases, the operating margin tends to decrease.

These correlation coefficients provide valuable insights for decision-making. For instance, strong positive correlations between **"Units Sold"** and **"Total Sales**," as well as between **"Total Sales"** and **"Operating Profit**," imply that increasing sales volumes can lead to higher revenues and profits. However, the negative correlation between **"Operating Profit"** and **"Operating Margin"** suggests that while profits rise, the profit margin tends to decrease.
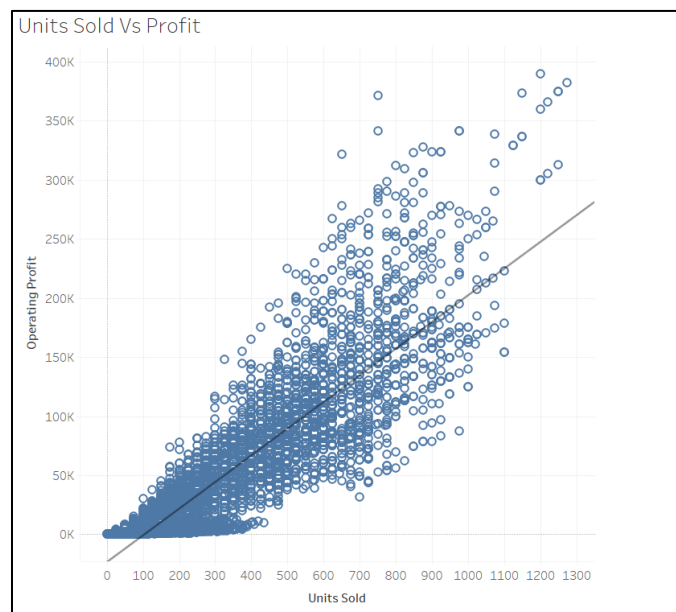
Above graph also suggests that there is linear

relationship between these 2 variables

4.1 Hypothesis Testing

Null Hypothesis : Assuming the Average Profit margin of Instore Equals Online Sales method

Alternate Hypothesis : Average Profit margin of Instore does not equals Online Sales method

To test the hypothesis, we considered the Z-test for 2 sample means

z-Test: Two Sample for Means

	Instore	*Online*
Mean	77309.51949	58318.93316
Known Variance	59953.29	65615.27
Observations	1103	1147
Hypothesized Mean Difference	0	
z	1797.970989	
P(Z<=z) one-tail	0	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0	
z Critical two-tail	1.959963985	

1. **Mean Comparison:**

   - The mean sales for the "Instore" channel is approximately $77,309.52, while the mean sales for the "Online" channel is approximately $58,318.93.

   - This suggests that, on average, the "Instore" channel has higher sales compared to the "Online" channel.

2. **Known Variances:**

   - The known variance for "Instore" is approximately 59,953.29, and for "Online," it is approximately 65,615.27. This information is used to calculate the z-test statistic.

3. **Sample Sizes:**

   - The sample size for "Instore" is 1103, and for "Online," it is 1147. The samples are of sufficient size for a statistical comparison.

4. **Hypothesized Mean Difference:**

   - The hypothesized mean difference is set to 0, indicating that the test is comparing whether there is a significant difference in means between the two channels.

5. **z-Test Statistic:**

   - The calculated z-test statistic is approximately 1797.97. This is a very large z-value, indicating a significant difference between the means.

6. **One-Tail Test:**

   - The p-value for a one-tailed test is very close to 0 (p(Z<=z) one-tail = 0), indicating strong evidence that the means are different in a specific direction (either "Instore" is significantly higher or "Online" is significantly higher).

7. **Two-Tail Test:**

   - The p-value for a two-tailed test is also very close to 0 (p(Z<=z) two-tail = 0), suggesting that there is a highly significant difference in means, regardless of the direction.

8. **z Critical Values:**

   - The critical z-value for a one-tail test is approximately 1.6449, and for a two-tail test, it's approximately 1.95996. Since the calculated z-value is significantly higher, it falls into the tail of the distribution, indicating a highly significant result.

In summary, the z-test results the null hypothesis can be rejected as Z value (**1797.97098940563**) is greater than the Z critical value(**1.959963985**)

4.2 Regression Analysis

Converted Sales Method to dummy variables
Instore-1
Outlet-2
Online -3

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.970648761							
R Square	0.942159017							
Adjusted R Square	0.942141025							
Standard Error	34136.32638							
Observations	9648							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	1.83054E+14	6.1018E+13	52362.99689	0			
Residual	9644	1.1238E+13	1165288779					
Total	9647	1.94292E+14						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	111593.0651	1761.673956	63.34490254	0	108139.8142	115046.316	108139.8142	115046.316
Operating Profit	2.421206439	0.006861825	352.85168	0	2.407755821	2.434657057	2.407755821	2.434657057
Operating Margin	-256431.0759	4000.800428	-64.09494314	0	-264273.4849	-248588.6669	-264273.4849	-248588.6669
Sales Method	2922.065485	532.4288592	5.488180129	4.16379E-08	1878.393112	3965.737859	1878.393112	3965.737859

Dependent Variable :Total Sales
Independent variables : Operating Profit, Operating Margin, Sales method

a. Regression Statistics:
   - Multiple R: 0.970648761
   - R Square: 0.942159017
   - Adjusted R Square: 0.942141025
   - Standard Error: 34136.32638
   - Observations: 9648

These statistics describe the goodness of fit of the regression model to the data. The R-squared value (R Square) indicates that approximately 94.22% of the variance in the dependent variable can be explained by the independent variables in the model.

b. ANOVA (Analysis of Variance):

  i.   Regression:
  ii.   df (degrees of freedom): 3
  iii. SS (Sum of Squares): 1.83054E+14
  iv.   MS (Mean Square): 6.1018E+13
  v.   F (F-statistic): 52362.99689
  vi.   Significance F: 0 (indicating that the regression model is significant)
  vii. Residual:
  viii. df: 9644
  ix.   SS: 1.1238E+13
  x.   MS: 1165288779
  xi.   Total:
  xii. df: 9647
  xiii. SS: 1.94292E+14

The ANOVA table shows the analysis of variance results for the regression model. The "Regression" section represents the variance explained by the model, while the "Residual" section represents the unexplained variance.

c. Coefficients:

  i.   Intercept: 111593.065061816
  ii.   Coefficient: 111593.0651
  iii.   Standard Error: 1761.673956
  iv.   t Stat: 63.34490254
  v.   P-value: 0 (indicating that the intercept is statistically significant)
  vi.   Confidence Interval (Lower 95% and Upper 95%): 108139.8142 and 115046.316
  vii.   Operating Profit:
  viii.   Coefficient: 2.421206439
  ix.   Standard Error: 0.006861825
  x.   t Stat: 352.85168
  xi.   P-value: 0 (indicating that Operating Profit is statistically significant)
  xii.   Confidence Interval (Lower 95% and Upper 95%): 2.407755821 and 2.434657057
  xiii.   Operating Margin:
  xiv.   Coefficient: -256431.0759
  xv.   Standard Error: 4000.800428
  xvi.   t Stat: -64.09494314
  xvii.   P-value: 0 (indicating that Operating Margin is statistically significant)
  xviii. Confidence Interval (Lower 95% and Upper 95%): -264273.4849 and -248588.6669
  xix.   Sales Method:
  xx.   Coefficient: 2922.065485
  xxi.   Standard Error: 532.4288592
  xxii.   t Stat: 5.488180129
  xxiii. P-value: 4.16379E-08 (indicating that Sales Method is statistically significant)
  xxiv.   Confidence Interval (Lower 95% and Upper 95%): 1878.393112 and 3965.737859

These coefficients represent the relationship between the independent variables (Operating Profit, Operating Margin, and Sales Method) and the dependent variable. The "t Stat" and "P-value" help determine the statistical significance of these coefficients.

In summary, this regression analysis shows a strong relationship between the independent variables and the dependent variable, with all three independent variables being statistically significant in predicting the dependent variable. The model explains approximately 94.22% of the variance in the dependent variable.

4.3 Forecasting

The Forecasting values for Moving Average and Exponential Smoothing are somewhat similar except that Moving Average is slightly more accurate. Therefore, the values for Moving Average are shown.
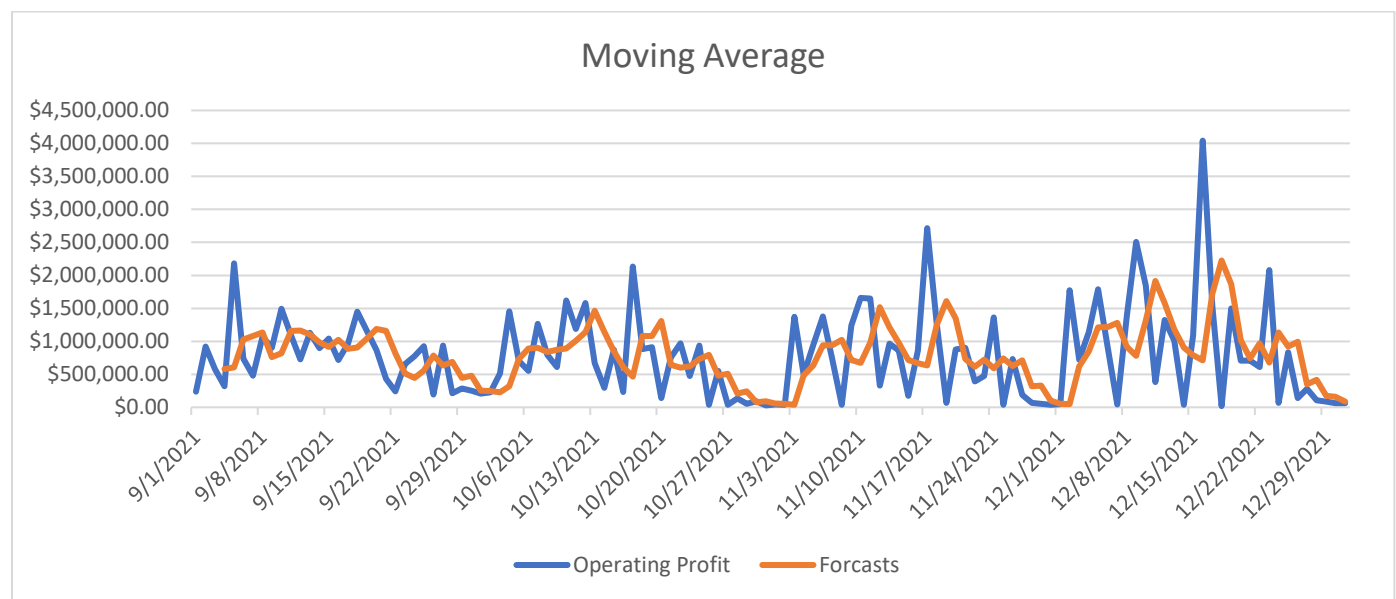
As presented in the figure above, there is a certain trend that may be followed. Every period, being a week, there are two points showing a minimum and two points showing a maximum. The values of importance are estimated below at:

Moving Absolute Error (MAE): $481,434.26

Moving Squared Error (MSE): $526,093,123,613.73

Moving Absolute Percentage Error (MAPE): 320.99%

The Errors as forecasted are far from the optimal value of zero. That being said, the Regression Model predicts total sales with 94.21% accuracy, operating profit being used as the X variable. However, the operating profit shifts from a very high actual value to a very low actual value at many chosen periods; explaining the irregular and inaccurate values. For the purposes of maintaining consistency and periodical relevance, the last three months of Adidas' operational profits were used to calculate the Moving Average forecast.

Time Stamps	Operating Profit	Forcasts	Forecast error	absolute error	Squared error	Percentage Error	Absolute Percentage Error
9/1/2021	$239,178.45						
9/2/2021	$919,486.17						
9/3/2021	$579,604.08						
9/4/2021	$318,289.16	$579,422.90	-$261,133.74	261133.74	$68,190,830,166.39	-82.042926	82.042926
9/5/2021	$2,183,657.04	$605,793.14	$1,577,863.90	1577863.903	$2,489,654,497,442.30	72.25786259	72.25786259
9/6/2021	$735,572.95	$1,027,183.43	-$291,610.48	291610.4767	$85,036,670,101.76	-39.64399135	39.64399135
9/7/2021	$481,595.19	$1,079,173.05	-$597,577.86	597577.86	$357,099,298,762.18	-124.0830208	124.0830208
9/8/2021	$1,073,167.69	$1,133,608.39	-$60,440.70	60440.70333	$3,653,078,619.43	-5.631990592	5.631990592
9/9/2021	$903,353.75	$763,445.28	$139,908.47	139908.4733	$19,574,380,910.46	15.48767283	15.48767283
9/10/2021	$1,494,377.32	$819,372.21	$675,005.11	675005.11	$455,631,898,526.11	45.16965702	45.16965702
9/11/2021	$1,093,212.00	$1,156,966.25	-$63,754.25	63754.25333	$4,064,604,818.09	-5.831828898	5.831828898
9/12/2021	$728,641.44	$1,163,647.69	-$435,006.25	435006.25	$189,230,437,539.06	-59.70100328	59.70100328
9/13/2021	$1,129,233.96	$1,105,410.25	$23,823.71	23823.70667	$567,568,999.34	2.109722831	2.109722831
9/14/2021	$895,999.13	$983,695.80	-$87,696.67	87696.67	$7,690,705,929.09	-9.78758428	9.78758428
9/15/2021	$1,042,223.81	$917,958.18	$124,265.63	124265.6333	$15,441,947,627.73	11.92312363	11.92312363
9/16/2021	$717,790.91	$1,022,485.63	-$304,694.72	304694.7233	$92,838,874,427.18	-42.44895263	42.44895263
9/17/2021	$963,125.21	$885,337.95	$77,787.26	77787.26	$6,050,857,818.31	8.076546974	8.076546974
9/18/2021	$1,450,900.09	$907,713.31	$543,186.78	543186.78	$295,051,877,966.77	37.43791759	37.43791759
9/19/2021	$1,156,446.16	$1,043,938.74	$112,507.42	112507.4233	$12,657,920,305.11	9.728721252	9.728721252
9/20/2021	$868,185.24	$1,190,157.15	-$321,971.91	321971.9133	$103,665,912,975.53	-37.08562395	37.08562395
9/21/2021	$429,974.08	$1,158,510.50	-$728,536.42	728536.4167	$530,765,310,409.51	-169.4372872	169.4372872
9/22/2021	$243,704.04	$818,201.83	-$574,497.79	574497.7867	$330,047,706,884.90	-235.7358486	235.7358486
9/23/2021	$656,681.23	$513,954.45	$142,726.78	142726.7767	$20,370,932,777.66	21.73456011	21.73456011
9/24/2021	$777,111.55	$443,453.12	$333,658.43	333658.4333	$111,327,950,134.46	42.93571925	42.93571925
9/25/2021	$927,190.99	$559,165.61	$368,025.38	368025.3833	$135,442,682,777.65	39.69251075	39.69251075
9/26/2021	$195,117.68	$786,994.59	-$591,876.91	591876.91	$350,318,276,591.15	-303.3435566	303.3435566
9/27/2021	$935,491.87	$633,140.07	$302,351.80	302351.7967	$91,416,608,947.56	32.32008811	32.32008811
9/28/2021	$211,820.61	$685,933.51	-$474,112.90	474112.9033	$224,783,045,107.16	-223.8275602	223.8275602
9/29/2021	$287,085.46	$447,476.72	-$160,391.26	160391.26	$25,725,356,284.39	-55.86882039	55.86882039
9/30/2021	$251,667.03	$478,132.65	-$226,465.62	226465.6167	$51,286,675,532.21	-89.98620783	89.98620783

As shown above, the actual operational profit values on 9/03/2021 to 9/04/2021 jump from $318,289 to $2,183,657. These values are what dictate the Absolute Errors to be higher than the actual values. Although it is true that when MAE, MSE and MAPE have high valuations, it does not necessarily mean that they are inaccurate. However, more information and detail is needed to conduct a more accurate forecast.

5.Conclusion

The project embarked on a comprehensive analysis of Adidas's US sales data to inform its potential entry into the North Korean market. This initiative underscores Adidas's commitment to data-driven decision-making and its adaptability in considering a challenging market like North Korea. The analysis involved answering critical questions, exploring the demographics of both the US and North Korea, understanding the provided dataset, conducting preliminary data analysis, employing statistical methods, and making sales forecasts.

The demographic analysis revealed the youthful population in North Korea, offering a potential market for Adidas's athletic products. Additionally, the low obesity rate in North Korea aligns with Adidas's focus on health and wellness, indicating a market opportunity for sports-related products.

The descriptive statistics of the Adidas US sales dataset provided essential insights into central tendencies, variability, and distribution of numerical variables. This information laid the groundwork for in-depth analysis and decision-making.

Furthermore, the correlation analysis uncovered significant relationships between various variables, offering valuable insights into how different factors impact sales and profitability. The hypothesis testing demonstrated a substantial difference in sales between the "Instore" and "Online" sales methods.

The forecasting results, while useful, highlighted the need for more detailed information to improve the accuracy of predictions. The Moving Average forecast, in particular, showed a significant margin of error, indicating that factors beyond the available dataset might influence sales and profitability.

In light of the project's findings, it is evident that entering the North Korean market is a complex endeavor, and more extensive research and data would be necessary to ensure accurate forecasts and informed decision-making. Adidas should consider further exploration of the North Korean market and gather more specific and up-to-date information to mitigate risks and increase the chances of success in this unique and challenging business environment.

6.Future Work

Future work on this project should focus on harnessing the power of both machine learning tools and other innovative technologies to propel Adidas's expansion into the North Korean market and refine its operations in the United States. Here's a roadmap for incorporating cutting-edge tools into the project:

Artificial Intelligence for Market Entry Strategy: Integrating AI systems to evaluate the complex political and economic landscape of North Korea is crucial. These systems can assess risks, predict market dynamics, and recommend strategic entry approaches.

IoT and Supply Chain Optimization: Leveraging the Internet of Things (IoT) to enhance supply chain management can optimize inventory, reduce costs, and ensure timely delivery. RFID and sensor technologies can provide real-time data on product movement and conditions.

Blockchain for Transparency: Implementing blockchain technology can ensure transparency and traceability in the supply chain. This will not only enhance trust but also meet regulatory requirements in both the US and North Korea.

Advanced Analytics for Customer Insights: Alongside machine learning, advanced analytics can extract deeper insights from customer data. Predictive analytics can anticipate trends, while prescriptive analytics can recommend precise actions to enhance customer experiences and product offerings.

Virtual Reality (VR) for Product Development: VR can revolutionize product design by creating virtual prototypes, conducting virtual focus groups, and allowing customers to 'try on' products virtually. This can expedite innovation and reduce costs.

Augmented Reality (AR) for Retail Experiences: AR can enhance in-store and online retail experiences. In physical stores, customers can use AR to access product information, while online shoppers can virtually 'try on' products. This fosters engagement and drives sales.

Chatbots and AI-Powered Customer Service: Deploying chatbots with natural language processing can improve customer service. They can handle routine inquiries and provide 24/7 support, enhancing customer satisfaction.

Quantum Computing for Advanced Data Processing: As quantum computing matures, it can be a game-changer for complex data analysis, enabling faster and more accurate decision-making.

Cybersecurity Enhancements: With the rising threat of cyberattacks, investing in cutting-edge cybersecurity solutions, including AI-based threat detection and blockchain-based security, is paramount to protect sensitive data.

Sustainability Initiatives: As sustainability gains importance, implementing tools to measure and optimize environmental impact can align with Adidas's mission and address growing consumer concerns.

Adidas's foray into North Korea presents unique challenges and opportunities. Future work on this project should not only integrate machine learning but also embrace these emerging technologies to ensure data-driven, innovative, and sustainable growth. This holistic approach will empower Adidas to navigate the intricacies of both the US and North Korean markets while setting new standards in the global sportswear industry.