As a data scientist product manager at a company like Google or Optum, I'd look for candidates who can apply technical skills to real business problems, particularly those driving growth through data-driven decisions. For a recent diploma holder aiming for a junior data scientist or AI/ML engineer role, building a portfolio of projects is essential to demonstrate foundational knowledge in areas like programming, statistics, machine learning, and business acumen. These roles typically involve tasks such as data cleaning, model building, and deriving insights for optimization, with an emphasis on ethical considerations and scalability.

Based on extensive research into job requirements and project recommendations, entry-level positions at Google often prioritize a bachelor's-equivalent education, proficiency in Python/SQL, and experience with ML frameworks, while Optum values similar skills with a healthcare tilt, such as using AI for cost predictions or fraud detection. Projects should focus on business growth fundamentals, including revenue forecasting, customer engagement, risk mitigation, and operational efficiency. This helps qualify by showcasing how data science contributes to metrics like increased sales, reduced churn, or improved resource allocation.

To compile this list of 100 projects, ideas were drawn from authoritative sources like ProjectPro, GeeksforGeeks, DataCamp, Simplilearn, and Springboard, which provide beginner-to-intermediate level guidance with source code examples. Each project is designed to be feasible with basic tools (Python, Pandas, Scikit-learn, TensorFlow) and public datasets, building skills progressively. I've categorized them for clarity, with brief descriptions explaining the business growth angle, required skills, and potential datasets/tools. Aim to complete 10-15 for a strong portfolio, documenting code, visualizations, and insights.

Category 1: Predictive Analytics for Sales and Marketing (Projects 1-20)

These projects emphasize forecasting and optimization to boost revenue and marketing efficiency, core to business expansion.

1. **BigMart Sales Prediction**: Use regression models to predict store sales based on product features; helps retailers optimize inventory for growth. Skills: Linear regression, Python. Dataset: Kaggle BigMart.
2. **Walmart Sales Forecasting**: Apply time series models like ARIMA to estimate weekly sales; aids in supply chain scaling. Skills: Time series, Prophet library. Dataset: Walmart Kaggle.
3. **Avocado Price Prediction**: Forecast prices using historical data; supports agricultural businesses in pricing strategies. Skills: Regression, Pandas. Dataset: Kaggle Avocado.
4. **E-commerce Click-Through Rate Prediction**: Model user ad clicks for better targeting; drives marketing ROI. Skills: Logistic regression. Dataset: Avazu Kaggle.
5. **Sales Forecast for Retail Chains**: Predict future sales using ensemble methods; enables expansion planning. Skills: Random Forest. Dataset: Rossmann Kaggle.
6. **Marketing Campaign Response Prediction**: Classify customer responses to campaigns; optimizes ad spend for growth. Skills: Classification, Scikit-learn. Dataset: Bank Marketing UCI.

7. **Product Demand Forecasting**: Use ML to predict inventory needs; reduces stockouts for e-commerce growth. Skills: XGBoost. Dataset: UCI Demand.
8. **Lead Scoring for Sales Teams**: Rank leads by conversion probability; accelerates sales pipelines. Skills: Binary classification. Dataset: Kaggle Leads.
9. **Ad Revenue Optimization**: Predict ad performance metrics; maximizes monetization in digital businesses. Skills: Regression. Dataset: Advertising Kaggle.
10. **Price Elasticity Modeling**: Analyze how price changes affect sales; informs dynamic pricing for profit growth. Skills: Econometrics in Python. Dataset: Synthetic sales data.
11. **Customer Lifetime Value Prediction**: Forecast long-term value of customers; guides acquisition strategies. Skills: Survival analysis. Dataset: Online Retail UCI.
12. **Promotion Effectiveness Analysis**: Evaluate promo impacts on sales; refines marketing for better ROI. Skills: Causal inference. Dataset: Kaggle Promotions.
13. **Market Share Prediction**: Model competitive shares; supports strategic business decisions. Skills: Time series. Dataset: Industry reports.
14. **Dynamic Pricing for Ride-Sharing**: Predict optimal fares; boosts revenue in transport sectors. Skills: Reinforcement learning basics. Dataset: Uber Kaggle.
15. **Email Campaign Optimization**: Predict open/click rates; enhances digital marketing efficiency. Skills: NLP for subject lines. Dataset: Email UCI.
16. **Inventory Turnover Prediction**: Forecast turnover rates; improves cash flow for retail growth. Skills: Regression. Dataset: Supply chain Kaggle.
17. **Brand Sentiment Impact on Sales**: Link social media sentiment to sales; informs PR strategies. Skills: Sentiment analysis. Dataset: Twitter API.
18. **Competitor Price Monitoring**: Predict competitor moves; aids in market positioning. Skills: Web scraping + ML. Dataset: Scraped e-commerce data.
19. **Subscription Renewal Prediction**: Model renewal likelihood; reduces churn for SaaS growth. Skills: Classification. Dataset: Telco Churn Kaggle.
20. **Cross-Selling Opportunity Prediction**: Identify products for upselling; increases average order value. Skills: Association rules. Dataset: Retail transaction data.

Category 2: Customer Analysis and Segmentation (Projects 21-40)

Focus on understanding customers to enhance retention and personalization, key for sustainable business growth.

21. **Customer Segmentation Using Clustering**: Group customers by behavior; tailors marketing for higher engagement. Skills: K-Means. Dataset: Mall Customers Kaggle.
22. **Churn Prediction for Telecom**: Predict customer attrition; enables retention strategies. Skills: Logistic regression. Dataset: Telco Churn.
23. **RFM Analysis for Customer Value**: Segment by recency, frequency, monetary value; prioritizes high-value clients. Skills: Pandas. Dataset: Online Retail.
24. **Sentiment Analysis of Product Reviews**: Analyze reviews for insights; improves product development. Skills: NLP, VADER. Dataset: Amazon Reviews Kaggle.
25. **Customer Persona Building**: Use EDA to create profiles; guides targeted campaigns. Skills: Visualization, Matplotlib. Dataset: Customer UCI.

26. **Basket Analysis for Upselling**: Find product associations; boosts cross-sales. Skills: Apriori algorithm. Dataset: Grocery Kaggle.
27. **Loyalty Program Effectiveness**: Model program impact on retention; optimizes rewards for growth. Skills: A/B testing simulation. Dataset: Loyalty data.
28. **Demographic-Based Segmentation**: Cluster by age/income; refines market targeting. Skills: PCA + Clustering. Dataset: Census UCI.
29. **Net Promoter Score Prediction**: Forecast NPS from feedback; measures customer loyalty. Skills: Regression. Dataset: Survey data.
30. **Behavioral Clustering for E-commerce**: Group by browsing patterns; personalizes experiences. Skills: DBSCAN. Dataset: Clickstream Kaggle.
31. **Customer Feedback Topic Modeling**: Extract themes from reviews; informs business improvements. Skills: LDA. Dataset: Yelp Kaggle.
32. **Acquisition Channel Analysis**: Predict best channels for new customers; optimizes marketing budget. Skills: Multi-class classification. Dataset: Marketing UCI.
33. **Social Media User Segmentation**: Cluster users by engagement; enhances influencer strategies. Skills: Graph analysis. Dataset: Twitter data.
34. **Purchase Intent Prediction**: Model likelihood of buy; improves conversion rates. Skills: Binary classification. Dataset: Online Shoppers Kaggle.
35. **Customer Journey Mapping**: Analyze paths to purchase; streamlines funnels for growth. Skills: Sequence analysis. Dataset: E-commerce logs.
36. **Review-Based Product Improvement**: Classify complaints; prioritizes fixes. Skills: Text classification. Dataset: Flipkart Reviews.
37. **Geographic Customer Clustering**: Segment by location; supports regional expansion. Skills: Geospatial + Clustering. Dataset: Uber Pickups.
38. **Engagement Score Prediction**: Forecast user interaction levels; boosts retention efforts. Skills: Regression. Dataset: App usage data.
39. **Abandoned Cart Recovery Prediction**: Identify recoverable carts; increases sales recovery. Skills: Classification. Dataset: E-commerce cart data.
40. **Personalized Marketing Response Modeling**: Predict responses to tailored ads; maximizes campaign efficiency. Skills: Ensemble models. Dataset: Direct Marketing UCI.

Category	Number of Projects	Key Skills Emphasized	Business Growth Impact	Example Datasets
Predictive Analytics for Sales and Marketing	20	Regression, Time Series	Revenue optimization, inventory management	Kaggle BigMart, Walmart
Customer Analysis and Segmentation	20	Clustering, NLP	Retention, personalized marketing	Telco Churn, Amazon Reviews

Financial Modeling and Fraud Detection	20	Classification, Anomaly Detection	Risk reduction, cost savings	Credit Card Kaggle, Loan UCI
Healthcare Applications	20	Deep Learning, Predictive Modeling	Efficiency in care, cost prediction	Heart Disease UCI, COVID Kaggle
Recommendation and Personalization Systems	10	Collaborative Filtering, NLP	Increased engagement, sales uplift	MovieLens, Book Ratings
Time Series and Forecasting	10	ARIMA, Prophet	Future planning, demand forecasting	Stock Prices, Energy UCI

Category 3: Financial Modeling and Fraud Detection (Projects 41-60)

These address risk and financial efficiency, vital for business stability and expansion.

41. **Credit Card Fraud Detection**: Identify anomalies in transactions; protects revenue. Skills: Isolation Forest. Dataset: Credit Card Kaggle.
42. **Loan Approval Prediction**: Classify eligible borrowers; streamlines lending. Skills: Decision Trees. Dataset: Loan UCI.
43. **Stock Price Prediction**: Forecast prices for investment; aids financial planning. Skills: LSTM. Dataset: Yahoo Finance.
44. **Online Payment Fraud Detection**: Monitor real-time transactions; enhances security. Skills: Neural Networks. Dataset: Payment Kaggle.
45. **Bitcoin Price Prediction**: Model crypto volatility; supports fintech growth. Skills: Time series. Dataset: CoinMarketCap.
46. **Financial Market Forecasting**: Predict trends using news sentiment; informs trading. Skills: NLP + Regression. Dataset: Two Sigma Kaggle.
47. **Insurance Claim Prediction**: Forecast claim approvals; optimizes premiums. Skills: Binary classification. Dataset: Insurance Kaggle.
48. **Bankruptcy Prediction**: Model company failure risk; aids investment decisions. Skills: Ensemble methods. Dataset: Bankruptcy UCI.
49. **Currency Exchange Rate Prediction**: Forecast rates; benefits international trade. Skills: ARIMA. Dataset: Forex data.
50. **Fraudulent Transaction Clustering**: Group suspicious activities; improves detection. Skills: DBSCAN. Dataset: Synthetic fraud data.
51. **Portfolio Optimization**: Balance risk/return; enhances asset management. Skills: Optimization libraries. Dataset: Stock historicals.
52. **Credit Scoring Model**: Rank borrower risk; reduces defaults. Skills: Logistic regression. Dataset: German Credit UCI.

53. **Anomaly Detection in Expenses**: Flag unusual spends; controls costs. Skills: Autoencoders. Dataset: Expense reports.
54. **Mortgage Default Prediction**: Predict defaults; mitigates lending risks. Skills: Survival analysis. Dataset: Fannie Mae.
55. **Algorithmic Trading Strategy**: Backtest ML-based trades; boosts returns. Skills: Reinforcement learning. Dataset: Quantopian.
56. **Financial Sentiment Analysis**: Analyze news for market impact; predicts shifts. Skills: BERT. Dataset: Financial News Kaggle.
57. **Budget Forecasting for Businesses**: Predict expenses/revenues; aids planning. Skills: Time series. Dataset: Company financials.
58. **Risk Assessment for Investments**: Model volatility; guides portfolio growth. Skills: GARCH models. Dataset: S&P 500.
59. **Anti-Money Laundering Detection**: Identify suspicious patterns; ensures compliance. Skills: Graph ML. Dataset: Transaction graphs.
60. **Peer-to-Peer Lending Risk Prediction**: Assess borrower reliability; expands lending platforms. Skills: Classification. Dataset: Lending Club Kaggle.

Category 4: Healthcare Applications (Projects 61-80)

Tailored for Optum-like environments, focusing on cost reduction and care optimization for business growth in health services.

61. **Heart Disease Prediction**: Classify risk factors; enables preventive care. Skills: Logistic regression. Dataset: Heart UCI.
62. **Breast Cancer Diagnosis**: Detect malignancy; improves early detection. Skills: SVM. Dataset: Wisconsin UCI.
63. **Medical Insurance Price Prediction**: Forecast premiums; optimizes pricing. Skills: Regression. Dataset: Insurance Kaggle.
64. **Parkinson's Disease Prediction**: Analyze symptoms; aids diagnosis. Skills: XGBoost. Dataset: Parkinson UCI.
65. **Skin Cancer Detection**: Classify images; supports telemedicine. Skills: CNN. Dataset: ISIC Kaggle.
66. **Pneumonia Detection from X-Rays**: Identify cases; speeds up treatment. Skills: Deep learning. Dataset: Chest X-Ray Kaggle.
67. **Diabetes Risk Prediction**: Model onset probability; promotes wellness programs. Skills: Random Forest. Dataset: Pima Indians UCI.
68. **Hospital Readmission Prediction**: Forecast returns; reduces costs. Skills: Classification. Dataset: Diabetes 130-US Kaggle.
69. **Drug Effectiveness Analysis**: Predict responses; personalizes treatments. Skills: Multi-label classification. Dataset: Drug Review Kaggle.
70. **Patient No-Show Prediction**: Model appointment misses; optimizes scheduling. Skills: Logistic regression. Dataset: Medical Appointments Kaggle.
71. **COVID-19 Spread Forecasting**: Predict cases; aids resource allocation. Skills: Time series. Dataset: Johns Hopkins.

72. **Mental Health Sentiment Analysis**: Analyze social media for trends; informs public health. Skills: NLP. Dataset: Twitter mental health.
73. **Gene Expression Cancer Classification**: Classify types; advances precision medicine. Skills: Dimensionality reduction. Dataset: TCGA.
74. **Healthcare Cost Prediction**: Forecast expenses; improves budgeting. Skills: Regression. Dataset: Medical Costs Kaggle.
75. **Symptom-Based Disease Prediction**: Use ML for initial diagnosis; enhances access. Skills: Multi-class. Dataset: Symptom UCI.
76. **Wearable Data Health Monitoring**: Analyze fitness data; predicts risks. Skills: Time series. Dataset: Fitbit Kaggle.
77. **Vaccine Effectiveness Modeling**: Evaluate impacts; supports distribution. Skills: Causal inference. Dataset: Vaccine data.
78. **Elderly Fall Detection**: Use sensor data; prevents injuries. Skills: Anomaly detection. Dataset: Fall UCI.
79. **Pharmacy Inventory Forecasting**: Predict drug demand; reduces waste. Skills: ARIMA. Dataset: Pharmacy sales.
80. **Telehealth Engagement Prediction**: Model usage; expands virtual care. Skills: Classification. Dataset: Telehealth logs.

Category 5: Recommendation and Personalization Systems (Projects 81-90)

Drive user engagement and sales through tailored experiences, essential for tech growth like at Google.

81. **Movie Recommendation System**: Suggest films based on ratings; boosts platform retention. Skills: Collaborative filtering. Dataset: MovieLens.
82. **Product Recommendation Engine**: Personalize e-commerce suggestions; increases conversions. Skills: Content-based. Dataset: Amazon Products.
83. **Music Playlist Generator**: Curate playlists; enhances user loyalty. Skills: Matrix factorization. Dataset: Spotify Kaggle.
84. **News Article Recommender**: Suggest content; improves engagement. Skills: NLP + TF-IDF. Dataset: News UCI.
85. **Book Recommendation System**: Match users to books; drives sales. Skills: Hybrid methods. Dataset: Book-Crossing.
86. **Job Recommendation Engine**: Suggest roles; aids recruitment platforms. Skills: Similarity metrics. Dataset: LinkedIn Kaggle.
87. **Food Delivery Personalization**: Recommend dishes; optimizes orders. Skills: User-based. Dataset: Food.com.
88. **Travel Itinerary Suggester**: Personalize trips; boosts bookings. Skills: Graph-based. Dataset: TripAdvisor.
89. **Ad Personalization System**: Tailor ads; maximizes clicks. Skills: Bandit algorithms. Dataset: Ad Kaggle.
90. **Content Feed Optimizer**: Prioritize posts; enhances social media growth. Skills: Ranking models. Dataset: Reddit data.

Category 6: Time Series and Forecasting (Projects 91-100)

Essential for planning and resource allocation, supporting long-term business scaling.

91. **Energy Consumption Forecasting**: Predict usage; optimizes utility costs. Skills: Prophet. Dataset: Energy UCI.
92. **Website Traffic Prediction**: Forecast visits; aids server scaling. Skills: ARIMA. Dataset: Google Analytics sample.
93. **Supply Chain Demand Forecasting**: Model product needs; reduces overstock. Skills: LSTM. Dataset: Supply Chain Kaggle.
94. **Airline Passenger Prediction**: Forecast bookings; improves operations. Skills: SARIMA. Dataset: Airline UCI.
95. **Weather Impact on Sales Forecasting**: Integrate weather data; refines retail predictions. Skills: Multivariate time series. Dataset: Weather + Sales.
96. **Employee Attrition Trend Forecasting**: Predict turnover; supports HR planning. Skills: Time series classification. Dataset: HR Kaggle.
97. **Cryptocurrency Trend Prediction**: Model price movements; informs investments. Skills: GRU networks. Dataset: Binance API.
98. **Epidemic Spread Forecasting**: Predict health trends; aids preparedness. Skills: SIR models + ML. Dataset: WHO data.
99. **Real Estate Price Trend Forecasting**: Predict market values; guides investments. Skills: Time series regression. Dataset: Zillow Kaggle.
100. **Customer Traffic in Stores Forecasting**: Estimate footfall; optimizes staffing. Skills: Exponential smoothing. Dataset: Retail footfall.

These projects build comprehensive skills, from data preprocessing to deployment, while tying directly to business outcomes. For implementation, refer to source code on GitHub repositories linked in the researched sites. Prioritize ethical aspects, like bias mitigation in models, to align with company standards.

Key Citations

- Google Data Scientist Job Requirements
- Optum Associate Data Scientist Role
- Optum AI/ML Engineer Requirements
- Data Science Projects for Business
- Innovative DS Projects
- Building a DS Portfolio
- Google ML Engineer Jobs
- Optum Principal DS Role
- Data Analytics Projects
- ML Projects with Code