Customer Segmentation and Clustering Analysis Presentation

Slide 1: Title Slide

- **Title**: Customer Analytics in Industry: Segmentation and Clustering
- **Subtitle**: Part 1 - Demographic-Based Customer Segmentation
- **Authors**: Vipul Sunil Patil, Vishal Pal
- **Institution**: Indian Institute of Technology, Gandhinagar
- **Supervisor**: Prof. Marcos Inácio Severo de Almeida
- **Visual**: Background image of a segmented pie chart or network of connected nodes

Speech:

Good morning, everyone! Thank you for joining us today. I'm Vipul Patil, alongside Vishal Pal, and we're excited to present our work on *Customer Analytics in the Industry: Segmentation and Clustering*. We're students at the Indian Institute of Technology, Gandhinagar, guided by Prof. Marcos Inácio Severo de Almeida. This is Part 1 of our project, where we use demographic data to segment customers into meaningful groups for smarter marketing. Picture this as a roadmap to understanding who your customers are—today, we'll uncover their demographic profiles, and in Part 2, we'll explore their buying habits. Let's get started!

Slide 2: Objectives

- **Title**: Objectives of the Study
- **Bullet Points**:
  - Segment customers using demographic features via clustering
  - Improve segmentation clarity with PCA dimensionality reduction
  - Deliver actionable insights for personalized marketing
  - Align findings with the STP framework and industry practices
- **Visual**: Icons (e.g., puzzle pieces for segmentation, magnifying glass for insights)

Speech:

Our journey begins with our objectives, which are the compass for this study. We set out to: (1) segment customers using demographic traits like age and income through clustering techniques; (2) sharpen those segments with Principal Component Analysis, or PCA, to simplify the data; (3) deliver clear, actionable insights for personalized marketing; and (4) align our work with the STP framework—Segmentation, Targeting, Positioning—a gold standard in industry practices. These goals ensure our analysis isn't just numbers on a page but a tool for real-world impact. Next, let's see how we made this happen.

Slide 3: Methodology Overview

- **Title**: Methodology
- **Bullet Points**:
  - Data: "segmentation data.csv" (2000 customers, 7 features)
  - Preprocessing: Feature standardization
  - Clustering: Hierarchical and K-means (optimal clusters via elbow method)
  - Dimensionality Reduction: PCA (3 components, ~80% variance)
  - Analysis: Cluster profiling and PCA interpretation
  - Visualizations: Heatmaps, dendrograms, elbow plots, PCA scatter plots
- **Visual**: Flowchart of methodology steps

Speech:

Here's our game plan, laid out in this flowchart. We started with the *segmentation data.csv* dataset—2000 customers, each with seven demographic features like Age, Income, and Education. These features were on different scales, so we standardized them—imagine converting everything to a common language so no feature overshadows another. Then, we used two clustering methods: hierarchical clustering to explore natural groupings, and K-means clustering to finalize them, guided by the elbow method. We also applied PCA to boil down our seven features into three key components, keeping about 80% of the data's story intact. Finally, we analyzed these clusters with tools like heatmaps and scatter plots, which you'll see soon. This structured approach ensures our findings are solid and easy to follow.

Slide 4: Data Exploration

- **Title**: Data Exploration
- **Bullet Points**:
  - Dataset: 2000 customers, 7 integer features
  - No duplicates detected
  - Mean Age: 35.91 years, Mean Income: $120,954
  - Correlation Heatmap: Strong links between Income, Occupation, Settlement Size
- **Visual**: Correlation heatmap

Speech:

Before clustering, we dug into the data to know what we're working with. Our dataset has 2000 unique customers—no duplicates—and seven integer features. On average, customers are 35.91 years old with an income of $120,954, but there's a wide range, as shown in the stats. Now, look at this correlation heatmap—it's like a relationship map. The colors show how features connect: red means a strong positive link, blue means negative, and white is neutral. See the dark red between Income and Occupation? That's a 0.68 correlation—higher-paying jobs often mean higher income. Occupation and Settlement Size, at 0.57, suggest professionals flock to bigger cities. Age and Income, though, have a weaker link at 0.17, hinting they'll play different roles in clustering. This heatmap was our first clue to how customers might group.

Slide 5: Hierarchical Clustering

- **Title**: Hierarchical Clustering
- **Bullet Points**:
    - Method: Ward's linkage
    - Dendrogram indicates 4 clusters
    - Truncated dendrogram for clarity
- **Visual**: Dendrogram

Speech:

Now, let's group those customers. We started with hierarchical clustering using Ward's method, which keeps clusters tight by minimizing variance. The result? This dendrogram—a tree-like diagram. Each vertical line is a customer, and horizontal lines show where they merge into clusters. The taller the line, the bigger the difference between groups. Look at the top: four main branches emerge, suggesting four clusters. We truncated it here to focus on the last five merges for clarity. Notice that one branch—later our Well-Off segment—stands tall and separate, hinting it's distinct, maybe older or wealthier customers. This dendrogram gave us a hunch of four clusters, which we'll test next with K-means.

Slide 6: K-means Clustering

- **Title**: K-means Clustering
- **Bullet Points**:
    - Optimal clusters: 4 (confirmed by elbow method)
    - Elbow plot shows WCSS stabilizes at 4 clusters
    - Clusters named: Well-Off, Career_Focused, Fewer_Opportunities, Standard
- **Visual**: Elbow plot

Speech:

To confirm our clusters, we turned to K-means clustering, which assigns customers to groups based on similarity. But how many groups? Enter the elbow method. We ran K-means from 1 to 10 clusters and plotted the Within-Cluster Sum of Squares—WCSS—basically, how spread out points are within each cluster. On this elbow plot, WCSS drops fast at first, then slows around four clusters—see that bend, like an elbow? That's where adding more clusters doesn't help much, so we picked four. We named them Well-Off, Career_Focused, Fewer_Opportunities, and Standard, based on their traits. This plot is our proof that four is the sweet spot—clear and manageable.

Slide 7: Cluster Profiles

- **Title**: Cluster Profiles
- **Table**:

Segment	Age	Income	Education	Occupation	Settlement Size
Well-Off	49	$164,420	2.01	1.28	1.25
Career_Focused	31	$102,860	1.11	0.54	0.31
Fewer_Opportunities	29	$99,008	0.42	0.43	0.37
Standard	37	$143,536	0.86	1.28	1.43

- **Visual**: Heatmap of mean feature values by cluster

Speech:

Meet our four segments in this table and heatmap. The table lists averages: Well-Off customers are 49 years old with $164,420 income, high education (2.01), and live in larger settlements (1.25)—think urban elites. Career_Focused are younger at 31, earning $102,860, with lower education and smaller towns. Fewer_Opportunities, at 29, have $99,008 income and the least education (0.42), while Standard, at 37, balance out with $143,536 and higher settlement sizes. The heatmap visualizes this: darker colors mean higher values. Well-Off glows dark across Age, Income, and Education, while Fewer_Opportunities stays light. This side-by-side view—table for numbers, heatmap for patterns—shows how unique each segment is.

Slide 8: Principal Component Analysis (PCA)

- **Title**: Principal Component Analysis
- **Bullet Points**:
  - 3 components retain ~80% variance
  - Component 1: Career and socioeconomic status
  - Component 2: Education and lifestyle
  - Component 3: Age and experience
- **Visual**: PCA loadings heatmap

Speech:

With seven features, visualizing clusters gets tricky, so we used PCA to simplify. PCA condenses our data into three components that capture ~80% of its variance—think of it as summarizing a book into key chapters. This heatmap shows loadings: how much each feature shapes a component. Component 1, with dark reds for Income (0.62) and Occupation (0.57), tracks career and wealth. Component 2 ties to Education (0.65) and Marital Status, reflecting lifestyle. Component 3 is all about Age (0.82), showing experience. Red means a strong positive pull, blue a negative one. PCA cuts through the noise, letting us plot clusters clearly, as you'll see next.

Slide 9: PCA Scatter Plot

- **Title**: Clusters in PCA Space
- **Bullet Points**:
  - Well-Off and Career_Focused distinctly separated
  - Standard and Fewer_Opportunities show slight overlap
  - PCA improves cluster visualization
- **Visual**: Scatter plot of PCA components with cluster labels

Speech:

Here's the payoff: our clusters in PCA space. This scatter plot uses Component 1 (x-axis, socioeconomic status) and Component 2 (y-axis, lifestyle). Each dot is a customer, colored by cluster—green for Well-Off, purple for Career_Focused, blue for Fewer_Opportunities, red for Standard. Black crosses mark cluster centers. Well-Off sits far right—high income and occupation—while Career_Focused is lower, less educated. Standard and Fewer_Opportunities overlap slightly near the center, sharing moderate traits, but PCA still separates them better than raw data. This plot proves our clusters aren't just numbers—they're visually distinct groups, ready for marketing.

Slide 10: Cluster Interpretation

- **Title**: Interpreting the Segments
- **Bullet Points**:
  - Well-Off: Affluent, older, urban professionals
  - Career_Focused: Young, ambitious, budget-conscious
  - Fewer_Opportunities: Youngest, low income, growth potential
  - Standard: Balanced, middle-aged, moderate income
- **Visual**: Icons (e.g., luxury car for Well-Off, briefcase for Career_Focused)

Speech:

Let's paint a picture of these segments. Well-Off are affluent, older urbanites—imagine seasoned professionals enjoying luxury goods. Career_Focused are young go-getters, early in careers, seeking value and convenience, like busy grad students. Fewer_Opportunities, the youngest, face tighter budgets but have potential—think entry-level workers. Standard are the everyman—middle-aged, stable, everyday shoppers. These icons—a luxury car for Well-Off, a briefcase for Career_Focused—bring them to life. Knowing who they are helps us talk to them effectively.

Slide 11: Marketing Implications

- **Title**: Marketing Implications
- **Bullet Points**:
  - Well-Off: Premium products, exclusive offers
  - Career_Focused: Convenience, budget-friendly options
  - Fewer_Opportunities: Value deals, loyalty programs
  - Standard: Mainstream products, mass-market campaigns
- **Visual**: Marketing strategy icons (e.g., diamond for premium, clock for convenience)

Speech:

So, what do we do with this? For Well-Off, offer premium products—think organic skincare or luxury tech—with exclusive deals. Career_Focused need convenience—meal kits or budget apps—pushed online. Fewer_Opportunities respond to value—discounts, loyalty perks—like bulk buys for tight budgets. Standard suit mass-market staples—family groceries, reliable gear—via broad campaigns. These icons—a diamond for premium, a clock for speed—match strategies to traits. Our analysis turns data into dollars by targeting smarter.

Slide 12: Limitations

- **Title**: Limitations
- **Bullet Points**:
  - Limited to demographic data, no behavioral insights
  - Static analysis, no temporal trends
  - PCA retains ~80% variance, some loss
- **Visual**: Caution icon

Speech:

No analysis is perfect. We used only demographic data—no buying habits or preferences—which limits the full customer story. It's also static—no tracking over time—so shifts in behavior are invisible. PCA keeps 80% of the variance, but 20% slips away, possibly blurring fine details. This caution icon reminds us: our insights are strong but not exhaustive. Future work can fill these gaps.

Slide 13: Recommendations & Future Work

- **Title**: Recommendations & Future Directions
- **Bullet Points**:
  - Well-Off: Launch premium lines
  - Career_Focused: Offer convenient products
  - Fewer_Opportunities: Provide bulk discounts
  - Standard: Target with mid-range campaigns
  - Future: Add purchase data, explore psychographics
- **Visual**: Roadmap graphic

Speech:

Our roadmap forward: For Well-Off, launch premium lines—high-end fashion or gadgets. Career_Focused get convenient options—ready meals, quick tech. Fewer_Opportunities need bulk deals or rewards to build loyalty. Standard thrive with mid-range, mass-appeal products. Next, Part 2 adds purchase data for richer insights, and psychographics—like hobbies—could deepen profiles. This plan blends now and next, driving value.

Slide 14: Conclusion

- **Title**: Conclusion
- **Bullet Points**:
  - Four segments identified: Well-Off, Career_Focused, Fewer_Opportunities, Standard
  - PCA improved cluster clarity
  - Insights support targeted marketing
  - Leads into Part 2: Purchase Analytics
- **Visual**: Summary graphic or thank-you image

Speech:

To sum up, we've carved out four segments—Well-Off, Career_Focused, Fewer_Opportunities, Standard—using clustering and PCA. PCA clarified these groups, making them actionable for marketing. Our findings fuel targeted strategies and lead into Part 2's purchase analysis. Thank you for listening—we're excited to hear your thoughts or questions!